蛋白质工程原理与技术

刘贤锡 主编

山东大学出版社
Shandong University Press

蛋白质工程原理与技术

主编　刘贤锡

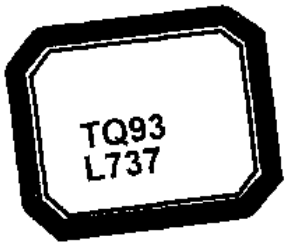
山东大学出版社

图书在版编目（CIP）数据

蛋白质工程原理与技术/刘贤锡等主编．—济南：山东大学出版社，2002.8
ISBN 7-5607-2465-5

Ⅰ．蛋…
Ⅱ．刘…
Ⅲ．蛋白质-化学加工
Ⅳ．TQ93

中国版本图书馆 CIP 数据核字（2002）第 053725 号



山东大学出版社出版发行
（山东省济南市山大南路 27 号　邮政编码：250100）
山东省新华书店经销
日照市黄海印刷厂印刷
787×1092 毫米　1/16　11.25 印张　257 千字
2002 年 9 月第 1 版　2002 年 9 月第 1 次印刷
印数：1—2300 册
定价：21.80 元

版权所有，盗印必究
凡购本书，如有缺页、倒页、脱页，由本社发行部负责调换



主　编　刘贤锡

副主编　王守训　刘师莲　高　聆　刘传华

编　者　（按姓氏笔画）
王守训　王　伟　王晓明
卞继峰　刘师莲　刘传华
刘贤锡　胡海燕　高　聆

主　审　于修平

前　言

蛋白质工程是继基因工程之后发展起来的一门新学科，它诞生于20世纪80年代之初，在近20年的时间内发展非常迅速。它是当今三大生物技术的领头羊，是新技术革命的重要组成部分，是基础理论研究向产业化转化的典范。蛋白质工程的原理及技术已渗透到医药学、农牧林科学及轻工业等多种领域，对这些领域的发展起着重要推动作用。

蛋白质工程是通过对蛋白质结构与功能关系的了解，借助于生物信息学的知识和手段，利用基因定点诱变和基因重组等技术，特异性地改造蛋白质的结构基因，产生具有新的特性的蛋白质的技术。其创造性成就在于开创了按照人类意愿设计制造符合人类需要的蛋白质的新时期。在近20年的实践过程中，无论是在基础理论研究领域，还是在生产实际应用方面，蛋白质工程研究均已获得了惊人的成就，使得生命科学的研究发生了深刻的变化，并将对今后生命科学的研究起到重大促进作用。因此，作为生物学和医药科学研究者有必要了解和掌握蛋白质工程的原理、技术，以及新进展和新成就。而本书正是根据蛋白质工程学科的进展以及教学的需要所编写。

本书共十章，第一章简要介绍了蛋白质工程的基本理论、基本内容及其程序、应用和进展；第二章至第四章分别介绍了蛋白质结构与功能、基因重组及定点诱变技术的基本原理；第五章至第七章分别介绍了蛋白质工程在酶蛋白质、抗体以及蛋白质及多肽药物研究中的应用；第八章和第九章分别介绍了蛋白质分离纯化与结构分析常用技术的原理及其应用；由于在蛋白质工程中常需要生物信息学对蛋白质的结构和功能进行分析并为蛋白质的结构改造进行预测，所以，本书第十章，对生物信息学进行了简要介绍。

本书的宗旨是为生物学和医学高级人才的培养提供一本有益的参考书。从这一宗旨出发，编者对本书的内容进行了精心的设计，并在系统介绍各章主要内容的同时，重点介绍了该领域的新进展和新成就。由于我们经验不足，水平有限，编写时间仓促，错误和不当之处在所难免，敬请同行专家以及使用本书的师生和读者批评指正。

编　者

2002年4月于山东大学医学院

目　录

第一章　绪论 …… (1)

第一节　蛋白质工程概论 …… (1)
一、蛋白质工程的理论基础 …… (1)
二、蛋白质工程的研究内容 …… (3)
三、蛋白质工程的基本程序 …… (4)
第二节　蛋白质工程的应用 …… (5)
一、研究蛋白质结构与功能的关系 …… (5)
二、改变蛋白质的特性 …… (6)
三、生产蛋白质和多肽类活性物质 …… (6)
四、设计合成全新蛋白质 …… (8)
第三节　蛋白质工程展望 …… (8)

第二章　蛋白质的结构与功能 …… (10)

第一节　蛋白质的基本结构与功能 …… (10)
一、蛋白质的组成 …… (10)
二、蛋白质的一级结构 …… (11)
三、蛋白质一级结构与功能的关系 …… (11)
第二节　蛋白质的空间结构与功能 …… (13)
一、蛋白质的二级结构 …… (13)
二、超二级结构和结构域 …… (13)
三、蛋白质的三级结构 …… (14)
四、蛋白质的四级结构 …… (16)
五、蛋白质空间结构与功能的关系 …… (16)
六、蛋白质一蛋白质相互作用 …… (18)

第三章　基因重组技术 …… (19)

第一节　工具酶 …… (19)

一、限制性核酸内切酶 …… (19)
二、其他工具酶 …… (21)
第二节 载体 …… (22)
一、质粒 …… (23)
二、噬菌体载体 …… (25)
三、表达载体 …… (27)
第三节 基因克隆的基本程序 …… (29)
一、目的基因的制备 …… (29)
二、目的基因与载体重组 …… (30)
三、重组 DNA 导入宿主细胞 …… (31)
四、重组体的筛选与鉴定 …… (32)
第四节 真核基因在大肠杆菌中的表达 …… (34)
一、真核基因在大肠杆菌中表达的基本条件 …… (34)
二、真核基因在大肠杆菌中的表达方式 …… (34)

第四章 定点诱变与蛋白质工程 …… (36)

第一节 定点诱变 …… (36)
一、M13DNA 寡核苷酸诱变 …… (37)
二、PCR 定点诱变 …… (39)
三、盒式诱变 …… (40)
四、随机诱变 …… (41)
第二节 蛋白质工程 …… (42)
一、加入二硫键 …… (42)
二、将 Asn 和 Gln 转换成其他氨基酸 …… (44)
三、减少游离 Cys 残基的数目 …… (44)
四、增加酶的活性 …… (45)
五、改变酶的特异性 …… (45)

第五章 酶蛋白质工程 …… (47)

第一节 酶的分子结构与催化功能 …… (47)
一、酶的结构特点与功能 …… (47)
二、酶催化作用的机制 …… (50)
三、酶促反应动力学参数 …… (52)
第二节 蛋白酶工程 …… (54)
一、枯草杆菌蛋白酶的定点改造 …… (54)
二、组织型纤溶酶原激活剂的定点改造 …… (56)
第三节 葡萄糖异构酶工程 …… (58)
一、结构特点及生物学性质 …… (58)

二、定点改造葡萄糖异构酶 …… (59)
三、应用 …… (60)
第四节　酶蛋白的定向进化 …… (60)
一、酶的定向进化的原理与思路 …… (61)
二、定向进化的常用技术 …… (62)

第六章　抗体工程 …… (65)

第一节　概述 …… (65)
一、多克隆抗体 …… (66)
二、单克隆抗体 …… (66)
三、基因工程抗体 …… (67)
第二节　抗体及其基因结构与抗体多样性的分子基础 …… (67)
一、抗体基因的基本结构 …… (67)
二、可变区基因的重组及其机制 …… (70)
三、CDR3 的多样性 …… (72)
四、抗体多样性的来源 …… (73)
第三节　鼠单抗的人源化 …… (73)
一、鼠单抗恒定区的人源化 …… (74)
二、鼠单抗可变区的人源化 …… (74)
第四节　小分子抗体 …… (75)
一、Fab 段 …… (76)
二、Fv 和单链抗体 …… (76)
三、胞内抗体 …… (78)
四、抗体融合蛋白 …… (79)
第五节　抗体库技术 …… (80)
一、初期的抗体库技术 …… (80)
二、噬菌体抗体库技术 …… (81)

第七章　蛋白质和多肽药物工程 …… (83)

第一节　概述 …… (83)
一、蛋白质多肽药物的表达系统 …… (83)
二、蛋白质多肽药物研究现状 …… (85)
第二节　激素类药物 …… (86)
一、胰岛素 …… (86)
二、生长激素 …… (88)
三、促红细胞生成素 …… (89)
第三节　细胞因子类药物 …… (89)
一、干扰素 …… (90)

二、白细胞介素 …… (91)
三、集落刺激因子 …… (93)

第八章 蛋白质分离纯化技术 …… (95)

第一节 蛋白质分析分离中的电泳技术 …… (95)
一、概述 …… (95)
二、聚丙烯酰胺凝胶电泳 …… (97)
三、SDS 聚丙烯酰胺凝胶电泳 …… (101)
四、等电聚焦电泳 …… (102)
五、双向凝胶电泳 …… (104)
六、毛细管电泳 …… (105)
七、聚丙烯酰胺凝胶印渍转移电泳 …… (109)
第二节 色谱技术 …… (109)
一、色谱技术分类 …… (110)
二、离子交换层析 …… (110)
三、凝胶过滤层析 …… (112)
四、分配层析 …… (115)
五、亲和层析 …… (116)
第三节 高效液相色谱 …… (117)
一、高效液相色谱基本理论 …… (118)
二、高效液相色谱固定相、流动相 …… (120)
三、高效液相色谱仪构成 …… (121)
四、高效液相色谱分析方法 …… (124)
五、氨基酸、多肽和蛋白质的 HPLC 分析 …… (125)

第九章 蛋白质结构分析技术 …… (129)

第一节 蛋白质和多肽的氨基酸序列分析 …… (129)
一、N 端序列分析 …… (129)
二、C 端序列分析 …… (133)
第二节 生物质谱技术在蛋白质和多肽结构测定中的应用 …… (135)
一、质谱技术的基本原理 …… (135)
二、生物质谱仪 …… (136)
三、生物质谱在蛋白质和多肽结构测定中的应用 …… (139)
第三节 X 射线晶体衍射技术在蛋白质晶体结构测定中的应用 …… (142)
一、基本原理 …… (142)
二、X 射线衍射技术在蛋白质分析中的应用 …… (143)
第四节 核磁共振波谱技术在蛋白质结构测定中的应用 …… (144)
一、NMR 的基本原理 …… (144)

二、生物 NMR 常用的实验技术 …… (150)
三、NMR 在蛋白质研究中的应用 …… (151)

第十章　生物信息学 …… (154)

第一节　概 述 …… (154)
第二节　生物信息数据库 …… (155)
一、基因和基因组数据库 …… (155)
二、蛋白质数据库 …… (157)
三、功能数据库 …… (158)
四、序列比对和预测分析 …… (160)
第三节　生物信息学在蛋白质研究中的应用 …… (164)
一、蛋白质组 …… (164)
二、蛋白质结构 …… (164)
三、新药设计 …… (164)

主要参考文献 …… (166)

第一章 绪 论

蛋白质工程（protein engineering）研究在过去的20余年间发展迅速，已取得一批较好成果，并开始应用于医学、农业、轻工等各个领域，产生了较大的经济和社会效益。2000年6月26日，人类基因组的工作草图宣告完成，标志着人类跨进21世纪历史新纪元之际，生命科学也迎来了一个崭新的时代，即后基因组时代（post－genome era）。在后基因组时代中，生物学的中心任务是揭示基因组及其所包含的全部基因的功能，并在此基础上阐明生命体的遗传、进化、发育、生长、衰老、死亡的基本生物学规律，以及与人类健康和疾病相关的生物学问题。由于基因的功能最终总是通过其表达产物蛋白质来实现的，因此，在人类基因组测序之后进一步集中研究蛋白质的结构及功能，是揭示基因组功能，阐释生命体重要生命活动规律和生命现象本质的基本途径，也是阐释疾病发生与发展的分子机理并进而战胜疾病的重要途径。而研究蛋白质的结构和功能，并在此项研究基础上人工改造蛋白质的结构，并获得我们所需要的活性蛋白质，正是蛋白质工程的主要任务和目标。

第一节 蛋白质工程概论

蛋白质工程是在重组DNA技术应用于蛋白质结构与功能研究之后发展起来的一门新兴学科。所谓蛋白质工程，就是通过对蛋白质已知结构和功能的了解，借助计算机辅助设计，利用基因定点诱变等技术，特异性地对蛋白质结构基因进行改造，产生具有新的特性的蛋白质的技术，并由此深入研究蛋白质的结构与功能的关系。蛋白质工程是在遗传工程取得的成就的基础上，融合蛋白质结晶学、蛋白质动力学、计算机辅助设计和蛋白质化学等学科而迅速发展起来的一个新兴研究领域，它开创了按照人类意愿设计制造符合人类需要的蛋白质的新时期，因此，被誉为第二代遗传工程。蛋白质工程的出现，为认识和改造蛋白质分子提供了强有力的手段。

一、蛋白质工程的理论基础

酶的专一性强，在温和条件下能有效地催化化学反应的能力使酶的应用日益广泛。药品、化学、食品工业及分析服务行业是酶开发的重要领地。目前已知的酶约有8000多种，至少有2500种有可能应用，而目前国际上工业用酶约为50种，以吨量级出售的

仅约 20 余种，可见，酶的开发利用尚有较大潜力。妨碍酶开发利用的原因主要有：①生物材料中酶含量甚少，用传统酶蛋白的分离纯化成本太高；②酶蛋白分子结构的稳定性差，过酸、过碱、高温、氧化等因素均可破坏其结构，使其丧失生物活性，因而在工业加工条件下，酶的半衰期短，利用率较低；③酶催化活性的最适 pH 值及底物专一性的范围较窄，与工业应用的要求有较大差距。因此，要想扩大酶蛋白的开发利用，需要建立适当的方法，以改善酶蛋白的生物特性，使之适合于工业应用的需求。

随着基因工程理论和技术的发展，人们已经能克隆特异蛋白质的基因，并令其在适宜宿主菌中表达，使蛋白质的产量大大提高，因而降低蛋白质纯化成本的问题已基本解决。重要的问题是提高酶蛋白的稳定性，改良其生物学特性。

稳定蛋白质空间构象的主要因素是蛋白质分子众多基团间的相互作用，如肽键和侧链间的氢键、带有不同电荷的侧链间的静电作用、极性氨基酸残基侧链间的偶极作用、疏水残基间的疏水作用以及分子内的二硫键等。蛋白质结构与功能研究表明，上述稳定蛋白质空间构象的因素是由蛋白质一级结构中某一个或某一段氨基酸序列决定的，人工改变或修饰这些氨基酸残基，有可能增加蛋白质的稳定性，又不影响其生物学活性。如野生型枯草杆菌蛋白酶的 218 位 Asn 是酶活性部位有关的氨基酸残基，若将其变成 Ser，用 65℃的失活半衰期衡量，从 59 分钟增加到 223 分钟，酶的稳定性显著增加。再如氧化失活使枯草杆菌蛋白酶在工业上的应用受到严重限制。很早就有人证实枯草杆菌蛋白酶在少量 H_2O_2 作用下很快失活是由于埋藏在催化部位 Ser221 邻近的 Met222 氧化成硫氢化物的原因。1985 年 Wells 证明若用其他氨基酸取代 Met222，可以提高酶的氧化稳定性而又保持高催化活性。

以上以酶为例介绍了蛋白质工程的重要性。实际上，在多肽药物、抗体以及其他活性蛋白质的基础与应用研究中，同样存在着蛋白质结构的改造问题。如白细胞介素－2（IL－2）是一种免疫反应调节因子，在医学上具有广泛的用途。结构研究证明，IL－2 是由 133 个氨基酸残基组成的多肽，有 3 个半胱氨酸，1 个二硫键，而 125 位上的半胱氨酸处于游离状态，这可能与其结构的稳定性有关。在分离纯化 IL－2 的过程中，人们发现在产品纯化和恢复活性过程中，IL－2 多肽链中的三个半胱氨酸之间易发生二硫键错配，致使整个多肽活性降低。对此，科学家通过定点诱变将编码链上编码 125 位半胱氨酸的密码子 TGT 转换成 TCT 或 GCA，也就是把半胱氨酸转换成丝氨酸或丙氨酸，结果可以避免二硫键的错配，使产品 IL－2 的活性提高了 7 倍。可见，改造蛋白质的空间结构以改变蛋白质的某些生物学特性，在拓宽蛋白质的用途，推动生物技术产业化方面至关重要。而蛋白质精细结构的改造，只有利用蛋白质工程技术才能得以实现。

蛋白质工程的重要目标之一，是通过改造蛋白质的结构来提高其开发利用的价值。这就是说，蛋白质的结构改变了，其生物学功能不能变。人所共知，蛋白质的功能是由其结构所决定的，蛋白质的结构改变了，其生物学功能能保持不变吗？蛋白质结构与功能关系的研究表明，蛋白质功能部位的几个氨基酸残基决定着蛋白质的功能，但是这几个负责功能的氨基酸残基必须处于一个极其精密的空间状态下才能发挥功能。这就是说，只要能维持蛋白质结构域的必须氨基酸及空间构象不变，其功能域的生物学活性就会保持不变。变异研究已经证明，许多蛋白质分子中多个氨基酸被改变，但其生物学功

能不受影响或只受极少的影响。事实上，在许多个体中已发现一些变异蛋白，伴有氨基酸的插入、丢失、或替换，但蛋白质仍保持正常功能。如前所述，将枯草杆菌蛋白酶催化部位 Ser221 邻近的 Met222 置换，可以提高该酶蛋白的抗氧化性，又能保持酶的催化活性，就是一个典型的例子。随着蛋白质三维结构分析技术的不断发展，借助于电脑对蛋白质结构域与功能域信息的处理及设计，采用定点诱变技术从基因水平改变氨基酸编码顺序即可获得与天然蛋白质的理化性质相异而生物活性相似的突变蛋白质。

生物技术的兴起使得分子生物学的理论与工程实践紧密结合。20 世纪 70 年代初期，DNA 重组技术诞生，并成功地应用于基因操作，从而产生了基因工程。而基因工程的诞生，特别是 DNA 重组技术和基因定点诱变等技术的建立，使我们有可能从基因水平改造蛋白质分子中氨基酸的序列，为蛋白质工程的诞生，奠定了技术基础。所谓基因定点诱变，就是在 DNA 水平上，通过对蛋白质结构基因中某个或某些氨基酸编码顺序加以改变，从而改变蛋白质结构的技术。该技术不仅可用于改造天然蛋白质，而且可以用于确定蛋白质中每一个氨基酸在结构和功能上的作用，以收集有关氨基酸线形顺序与其空间构象和生物学活性之间的对应关系，为人工改造蛋白质提供理论依据。

总之，蛋白质结构和动力学研究是蛋白质结构与功能关系研究的重要手段，而改造蛋白质结构，则依靠基因工程，基因工程的发展从技术上提供了改变蛋白质个别氨基酸残基或肽段的手段，使结构改变已能在实验室实施，二者结合则产生了蛋白质工程。

二、蛋白质工程的研究内容

(一) 利用已知蛋白质一级结构的信息作为应用研究

蛋白质的结构决定蛋白质的功能和理化性质，而蛋白质功能区的某个或某些氨基酸残基可能在维持蛋白质的结构、功能、理化性质中起重要作用。因此，定点改变这些氨基酸序列，即可能改变蛋白质的功能和特性，使之更适合于工业化生产的要求。随着蛋白质结构与功能关系研究技术以及基因工程技术的发展，上述设想已不再是可想而不可及。通过蛋白质结构与功能研究，利用蛋白质构型信息的图像和程序分析，即可找出与蛋白质功能和特性密切相关的氨基酸序列，而采用定点诱变与盒式突变技术定向改变编码蛋白质的基因，即可改造蛋白质的结构，使它们的生物功能及理化特性得到改变。如前所述枯草杆菌蛋白酶之所以易被氧化失活，是由于催化部位的 Ser221 邻近的 Met222 易被氧化成硫氧化物，若以其他氨基酸取代 Met222，则可提高酶的氧化稳定性而又不影响其催化活性。这是结构分析和定点突变改造蛋白质的成功范例。枯草杆菌蛋白酶可作为洗涤剂的添加剂，但由于其只能水解 Phe 羧基所形成的肽键，底物作用范围过窄而限制了洗涤剂的高效性，若用带正电荷的 Lys 取代位于活性中心 166 位的 Gly，所获得的突变酶不仅能水解 Phe 羧基所形成的肽键，而且可以水解酸性氨基酸 Glu 所形成的肽键，使其底物作用范围拓宽，因而可能成为最高效的洗涤剂添加酶，这是定点诱变改变蛋白质生物学活性的成功例子。

（二）从混杂变异体库中筛选和选择具有一定结构—功能关系的蛋白质

蛋白质的变异是蛋白质工程中很重要的研究内容和手段。研究者可有目的地在特定位点上使蛋白质发生变异，然后研究其结构与功能的关系。如果有了一个混杂的变异体库，也可从中筛选出具有一定结构—功能关系的蛋白质。例如，有人把对热不稳定的酶的基因转移到嗜热生物体内，然后利用酶的某种标志（如对卡那霉素的抗性）筛选出对热稳定的酶，既保持着酶的原来性质，又增加了热稳定性。目前从非定点诱变产生的或特定条件诱发的变异库中筛选，已得到耐热的卡那霉素核苷酰基转移酶。这是从混杂变异体库中筛选改造蛋白质性质的一个成功例子。这一类研究的主要困难是变异体库的获得及筛选方法的建立。

（三）定量蛋白质结构与功能关系的研究

这是目前蛋白质工程研究的主体。这一类研究的课题很广泛，包括蛋白质三维结构模型的建立，配体结合和酶催化的性质，蛋白质折叠和稳定性研究，蛋白质变异等等。这类研究看起来偏重理论，但对于指导实际工程意义重大。

（四）根据已知结构—功能的关系人工改造蛋白质

蛋白质工程是研究蛋白质结构和定点改造蛋白质结构的一门学科，其创造性成就在于开创了按照人类意愿设计制造符合人类需要的蛋白质的新时期。根据蛋白质结构和功能的关系，采用基因定位诱变工程技术，人工改造蛋白质则是蛋白质工程的主要研究内容之一。其主要研究内容如下：①通过改变蛋白质的活性部位，提高其生物功效及独立工作的能力；②通过改变蛋白质的结构顺序，提高其在极端条件（如酸、碱、热等）下的稳定性；③通过改变蛋白质的结构顺序使其便于分离纯化。这项研究虽然难度较大，但随着人类基因组研究、后基因组研究的不断深入和完善，随着蛋白质工程研究技术和手段的不断发展，该项研究必将获得丰硕的成果。

三、蛋白质工程的基本程序

蛋白质工程是从 DNA 的水平改变基因入手，设计合成或改造蛋白质的技术。该技术综合运用蛋白质的三维结构，结构与功能的详细信息，重组 DNA 基因操作技术，用定点诱变基因的方法直接修饰改变或人工合成基因，有目的地按照设计来改变蛋白质分子中某一种氨基酸残基或结构区域，从而定向改变蛋白质的性质，使其成为具有人们所希望性能的新型蛋白质，或者创造自然界不存在的性质独特的蛋白质（见图 1—1）。

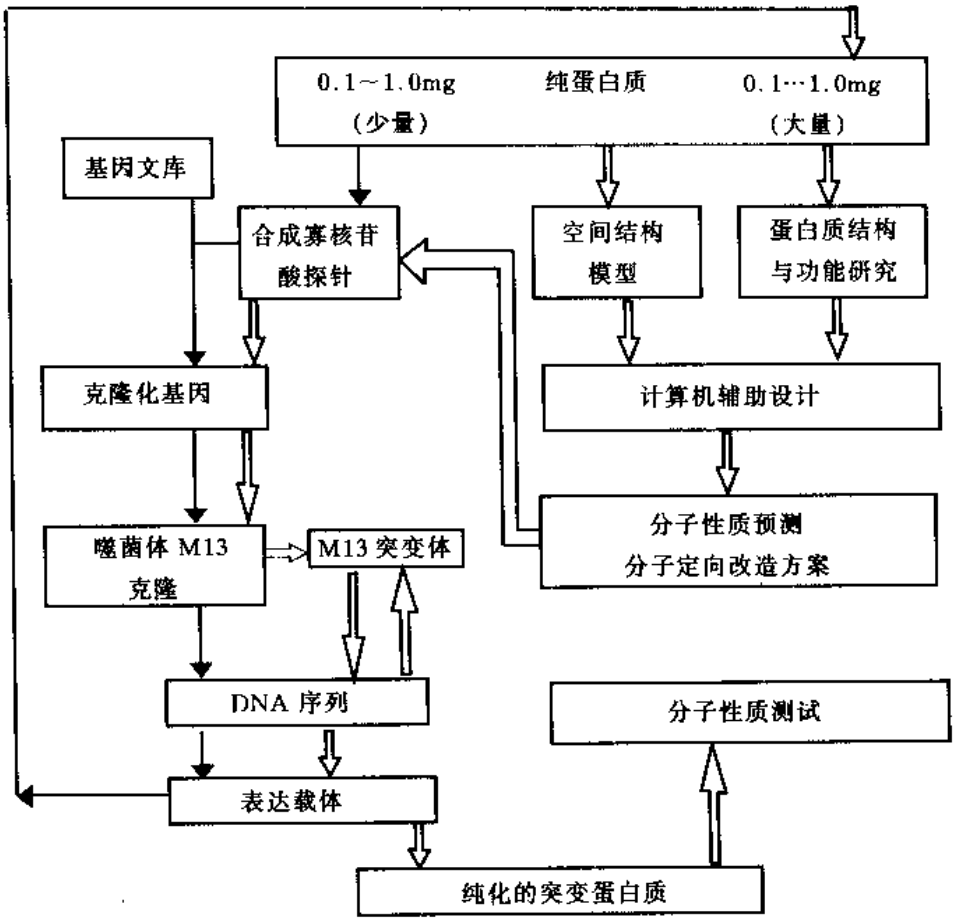


图 1－1 蛋白质工程的程序

首先分离纯化 0.1～1.0mg 纯蛋白质，测定其部分肽段的一段结构，根据编码原则合成相应同位素标记的寡核苷酸探针，以此从基因库中分离编码该蛋白质的克隆化基因，转入噬菌体 M13 系统，用双脱氧末端终止法进行 DNA 序列分析。通过表达载体获得较大量（0.1～1.0 克）该蛋白质用于空间结构测定及结构与功能研究，借助计算机提出分子预测性质及改造方案。通过寡核苷酸 M13 系统定点诱变并分离其突变体，引入表达载体生产并纯化多量突变性蛋白质，分析其性质指导进一步分子设计，以最终获得所预期性质的分子。

第二节 蛋白质工程的应用

蛋白质工程的研究层次已经深入到基因内部，而且所进行的一切改变都是理性的，有目的的和有根据的，其结果一般的说是可以预期的，而不是盲目的，因此，蛋白质工程从一开始就显示出无限的发展前景。在 20 余年的实践过程中，无论是在基础理论研究领域，还是在生产实际应用方面，蛋白质工程研究均已取得了惊人的成绩，使得生命科学的研究发生了深刻的变化。

一、研究蛋白质结构与功能的关系

研究蛋白质结构与功能的关系是按照人类的意愿改造蛋白质的特性、产业化生产活

性蛋白质、甚至设计和制造全新蛋白质的基础。而蛋白质工程则为研究蛋白质结构与功能的关系提供了必要的理论和技术，使我们有可能利用基因工程技术研究蛋白质的结构与功能的关系。

胰蛋白酶和弹性蛋白酶的酶原在激活时发生类似的构象变化，它们的三维结构和催化机制也很相似，但是，在对底物专一性上都表现出较大差异。晶体结构研究显示，二者与底物结合部位都有一个“口袋”，口袋边上的第 216 和 226 位氨基酸残基，胰蛋白酶均是甘氨酸（Gly 216，Gly226），弹性蛋白酶分别为缬氨酸（Val216）和苏氨酸（Thr226）。Val 和 Thr 的侧链比 Gly 的大，由于空间位阻关系，弹性蛋白酶的专一性底物只能是小分子的疏水氨基酸，胰蛋白酶的口袋则允许大分子氨基酸进入。由于胰蛋白酶口袋的底部第 189 位氨基酸残基是天冬氨酸（Asp189），能与带正电荷的氨基酸侧链基团静电结合，这决定了其只能专一性的水解精氨酸（Arg）和赖氨酸（Lys）羧基所形成的肽键。Craik 用定点诱变获得 Ala（丙氨酸）216，Ala226 和 Ala216＋Ala226 三种突变体，以研究氢原子被甲基取代所产生的空间效应对底物专一性以及酶的催化活性的影响。Craik 等将三个突变酶分别在猴肾细胞表达，并对产物进行了动力学分析，发现三种突变酶的活力均下降，但对 Arg 肽键和 Lys 肽键的水解速度的下降在不同突变酶中表现不同。例如突变酶 Ala216 对 Arg 肽键的催化活性高于对 Lys 肽键，这是因为 216 位引入甲基，使 414 位水分子因空间障碍不复存在，底物赖氨酰与酶的结合是通过该水分子与 189 位 Asp 形成的氢键网络，而底物精氨酰可以通过胍基直接与 Asp189 形成氢键，并不必需依赖于水分子，因此 Ala216 突变体对赖氨酰肽键催化活性的降低较对精氨酰肽键更为明显。与此相反，Ala226 位甲基由于空间障碍，影响了 414 位水分子的存在以及精氨酰通过该水分子与 Asp189 的氢键配对，从而大大降低了它对精氨酰肽键的水解速度。双突变酶 Ala216＋Ala226 由于 216，226 位两个甲基的取代使结合口缩小，而且由于局部构象的变化使结合状态的底物的取向可能与酶的活性部位不相匹配，因而催化活性最低。这项研究充分证明了酶活性部位空间位阻对酶催化功能的影响。这种蛋白质的精细结构改变对其功能影响的研究，只有通过蛋白质工程才能完成。

二、改变蛋白质的特性

蛋白质工程的建立，使人们有可能通过改造蛋白质的结构来改变蛋白质的生物学特性。如改变酶的催化活性、酶对底物的专一性、酶与配体的结合能力、酶在 pH 变化、温度变化以及溶剂系统变化条件下分子结构的稳定性等等。这些都可以通过蛋白质工程将酶分子中某个或某一段氨基酸残基更换、增加、删除或修饰来实现。有关这方面的例子本书中多有述及，此处不再赘述。

三、生产蛋白质和多肽类活性物质

（一）提高酶的产量和创建新型酶

天然酶的产量低、稳定性差，限制了酶的开发利用。DNA 重组技术对酶工业的渗透，导致了酶工业的迅速发展。目前已有多种酶实现了克隆化，并在适宜宿主菌中表达

成功，从而极大地提高了酶的产量。蛋白质工程技术的兴起，给创建新型酶、改变酶的生理特性提供了强有力的工具。如利用基因定点诱变技术，将 T4 溶菌酶的第 3 位异亮氨酸变成半胱氨酸，后者与第 97 位的半胱氨酸形成二硫键，从而稳定了两个结构域所形成的活性中心，使酶在高温下的稳定性大大提高。如果将该酶的第 51 位苏氨酸换成脯氨酸，可使酶的活性提高 25 倍。蛋白质工程技术使酶的产量增加，稳定性提高，从而促进了酶的开发利用。当前商品化的蛋白质工程酶，除了药用酶外，还有枯草杆菌蛋白酶、水解酶、酯酶、α-淀粉酶和凝乳酶等多种。

（二）设计和研制新型抗体

传统的抗体是经过抗原免疫动物获得的，称为多克隆抗体。1975 年，Kohler 及 Milstein 等通过杂交瘤技术制备了单克隆抗体（简称单抗），这是免疫学上划时代的伟大进展之一。单抗用于临床疾病治疗，效果不尽如人意，原因在于目前用于治疗的单抗多为小鼠单抗，对人而言是一种异体蛋白，会引起人产生人抗鼠抗体（human anti－mouse antibodies，HAMA），此时再使用小鼠单抗，HAMA 与小鼠单抗结合后，不但会中和单抗使之失效，还会产生有害的过敏反应。为解决这一问题，人们应用基因工程等手段进一步研究抗体的结构与功能的关系，并对抗体基因进行改造，制备出了第三代抗体——基因工程抗体（genetic engineering antibody，GeAb），使抗体的制备更加简化，并且拓宽了抗体的应用范围。

基因工程抗体技术主要包括两部分内容：①对已有的单克隆抗体进行改造，包括单抗的人源化、小分子抗体以及抗体融合蛋白的制备；②通过抗体库的构建，使得抗体不需抗原免疫即可筛选并克隆新的单抗。运用蛋白质工程技术对抗体进行改造，被称为抗体工程，主要包括嵌合抗体、重构抗体、单链抗体、单区抗体及新近出现的全套抗体的研究。

（三）设计和研制多肽及蛋白质类药物

随着蛋白质工程技术的建立和发展，医药生物高科技产品也迅速发展。当前，人们不仅可以利用基因克隆和表达技术，使可以应用于临床的微量蛋白质（如胰岛素、干扰素、某些酶类等）得以产业化生产，而且可以应用蛋白质工程“重新设计”的思想，以 DNA 重组技术为手段，对多肽和蛋白质类药物的结构加以改造，以期更好地解决蛋白质类药物的导向性问题。如把白介素 2 的基因同白喉毒素的第Ⅱ和第Ⅲ区段拼接，通过基因工程获得白介素 2－白喉毒素的杂合蛋白。该杂合蛋白可定向与 T 细胞表面的白介素 2 受体结合，进入细胞继而杀死 T 细胞，因而具有特异性的治疗 T 细胞白血病的作用，并可抑制器官移植的免疫排斥作用和治疗某些自身免疫性疾病。目前已能产业化生产并应用于临床治疗的基因工程药物有激素类及神经递质类药物、细胞因子类药物、酶类药物及凝血因子等不同类别的多种药物。基因工程药物是目前国内外发展较快的一个产业，同时也使人类疾病的防治水平得到了提高。

四、设计合成全新蛋白质

根据蛋白质结构与功能关系的研究结果，利用计算机对信息进行分析，通过实验，即可设计出一种预定空间结构的全新多肽和蛋白质分子，再利用基因重组、基因克隆和基因表达技术即可制造出具有全新化学性质、生物学活性的蛋白质。全新蛋白质的设计合成是一项十分复杂的工作，目前虽已有人报道已合成某种模拟蛋白质，但仍处于实验初级阶段。相信，随着蛋白质工程技术的发展，设计合成全新蛋白质的工作，也会获得突破性的进展。

第三节 蛋白质工程展望

蛋白质工程是20世纪80年代初诞生的一个新兴领域，其主要内容是将核酸研究与蛋白质研究相结合，通过有控制的基因修饰和改造，产生符合人们意愿的蛋白质。另一项更为新颖的战略是利用蛋白质组合规律及结构与功能关系方面的研究结果设计新的蛋白质。

蛋白质三维结构的研究起始于20世纪50年代，当时主要是利用X射线衍射的原理解析蛋白质晶体中的原子在空间的位置和排列。至今已有400多种蛋白质的三维晶体结构被搞清，为探讨蛋白质的结构与功能的关系提供了重要信息。然而，蛋白质晶体学所能提供的毕竟是一个静态的结构，更深一步的问题是从编码基因核苷酸序列翻译形成的或体外变异产生的线形多肽链是如何折叠成具有三维特征的天然结构？蛋白质在与其他物质（蛋白质、核酸或酶的底物等）相互作用时，其三维结构又是经历怎样的动态过程以发挥其活性？这是蛋白质动力学的重要研究课题，只有了解这些问题，才能预测基因水平的改造最终会对蛋白质的结构与功能产生何种后果，才能谈得上具有真正意义的分子设计。尽管人们目前在此领域的工作尚很少，但是，根据该项工作的原理及所需分析手段，对于蛋白质工程来说，是一种可行性研究工作。

现阶段蛋白质工程亟待解决的一个主要问题是发展克隆基因的表达方法，即找到理想的宿主表达系统。目前常用的表达系统有细菌表达系统、酵母表达系统、昆虫表达系统、哺乳动物细胞等，这些表达系统各有优缺点。细菌表达系统培养方法简单、增殖快、生产成本低，但缺乏真核细胞特有的加工后处理，而且在菌体中表达的外源蛋白，在原核细胞中不稳定，易被菌体蛋白酶破坏。酵母表达系统是一个重要的真核表达系统，既具有原核细胞的增殖快、操作简单、成本低等优点，又可以像其他真核细胞一样完成对蛋白质转录后的修饰，也能进行诱导性分泌，但它具有活化的蛋白水解酶类，可降解异体蛋白质，使产品产量降低。昆虫表达系统的优点是能较高水平地表达不同动物来源的基因，可以胞内表达，也可以进行分泌性表达，能够有效地进行蛋白质的翻译后加工。其主要缺点是不能连续合成重组蛋白，因为重组病毒感染昆虫细胞或昆虫幼虫4～5天后，细胞就裂解，幼虫即死亡。哺乳动物细胞是生产哺乳动物蛋白质最好的场所，其表达真核基因比以前几种表达系统更优越，能对蛋白质进行各种类型的加工和修饰，还能表达有功能的膜蛋白及分泌性蛋白。其缺点是组织细胞培养的技术要求高，培

养和筛选细胞株的周期长，成本较高。因此，选择和建立物美、价廉的表达体系，是蛋白质工程急需解决的问题。

（刘贤锡）

第二章　蛋白质的结构与功能

蛋白质（protein）是一类生物大分子，在细胞中含量丰富、种类繁多，约占细胞干重的70%以上。蛋白质在生物体内分布广泛，几乎存在于所有的器官和组织，是一切生命活动的物质基础。生物物种的现象、新陈代谢的类型和各种生命现象都是通过蛋白质来实现的。因此，蛋白质被称为“功能分子”，不同蛋白质的功能不同，蛋白质的功能由其结构所决定。本章主要介绍蛋白质的结构及其与功能的关系。

第一节　蛋白质的基本结构与功能

一、蛋白质的组成

蛋白质的组成单位是氨基酸，组成人体蛋白质的氨基酸有20种（见表2－1）。氨基酸借肽键连接形成多肽链。多肽链两端分别有一个游离的氨基和游离的羧基，前者称为氨基末端或N末端（amino terminal），后者称为羧基末端或C末端（carboxyl terminal）。由氨基酸借肽键相连组成的化合物称为肽，一般由10个以下氨基酸残基组成的肽称为寡肽（oligopeptide），由10个以上氨基酸残基组成的肽称为多肽（polypeptide）。

表2－1　氨基酸及其常用的代号

中文名称	英文名称	中文代号	三字代号	单字母代号
甘氨酸	glycine	甘	Gly	G
丙氨酸	alanine	丙	Ala	A
缬氨酸	valine	缬	Val	V
亮氨酸	leucine	亮	Leu	L
异亮氨酸	isoleucine	异亮	Ile	I
丝氨酸	serine	丝	Ser	S

中文名称	英文名称	中文代号	三字代号	单字母代号
苏氨酸	threonine	苏	Thr	T
天冬氨酸	aspartic acid	天冬	Asp	D
谷氨酸	glutamic acid	谷	Glu	E
组氨酸	histidine	组	His	H
羟赖氨酸	hydrolysine	羟	Hyl	
赖氨酸	lysine	赖	Lys	K
精氨酸	arginine	精	Arg	R
天冬酰胺	asparagine	天酰	Asn	N
谷氨酰胺	glutamine	谷酰	Gln	Q
半胱氨酸	cysteine	半胱	Cys	C
甲硫氨酸	methionine	甲硫	Met	M
苯丙氨酸	phenylalanine	苯（丙）	Phe	F
酪氨酸	tyrosine	酪	Tyr	Y
色氨酸	tryptophan	色	Trp	W
脯氨酸	proline	脯	Pro	P
羟脯氨酸	hydroproline	羟脯	Hyp	

二、蛋白质的一级结构

蛋白质的一级结构（primary structure）是指多肽链中氨基酸的排列顺序。通常由左至右，从氨基端开始，依次用氨基酸缩写符号向羧基端书写。蛋白质的一级结构是基因上的遗传信息所决定的。一级结构是蛋白质分子的基本结构，也是蛋白质空间结构的基础。通过一级结构的信息可以说明生物进化、遗传变异以及蛋白质功能的分子机理。蛋白质的一级结构可采用传统的测序方法来测定，也可应用质谱仪及从相应基因的一级结构来推测。

三、蛋白质一级结构与功能的关系

某些蛋白质的一级结构中有一些特定的序列，能够阻碍蛋白质表现出生物学功能，只有将其删除才能形成有功能的活性蛋白质。也有一些蛋白质因一级结构序列的差异可能影响蛋白质的正常功能而引起疾病。故蛋白质的一级结构与蛋白质的生物学功能密切相关。

（一）蛋白质的前体与活性蛋白

许多活性蛋白质都有一个不具有活性的前体，其英文字首用 pro 表示，中文一般译

成肽。前体需经专一性的蛋白水解酶切除一个或数个肽段，才能转变成具有生物活性的蛋白质。如胰脏中含有胰蛋白酶原、胰凝乳蛋白酶原、弹性蛋白酶原和羧肽酶原 A 和 B。当胰蛋白酶原被分泌进入十二指肠后，经肠激酶催化水解下一个六肽，从而使分子构象发生变形，形成酶的活性中心，转变成有活性的酶，这种现象称为酶原激活。胰蛋白酶原不仅能够被肠激酶激活，它还有自身催化激活作用，并能激活消化道中的其他蛋白酶原。

某些肽类及蛋白类激素原也需经过酶切激活才能转变成有生物活性的激素。例如人胰岛素初合成时是一条由 110 个氨基酸残基组成的前胰岛素原（preproinsulin）肽链，经内质网膜上的信号肽酶作用，剪切掉 N－端 24 个氨基酸残基的信号肽序列（signal sequence），形成含有 86 个氨基酸残基组成的胰岛素原（proinsulin），并在相应位置上形成 3 个二硫键。当其在高尔基器包装成为颗粒时，再经胰蛋白酶和类羧肽酶 B 的催化，切除中间 33 个氨基酸残基的 C 肽（连接肽）后，才能转变成具有生物活性的胰岛素。

（二）蛋白质一级结构与分子进化

生物进化不仅表现在生物的外部形态上，也反映在蛋白质结构上。因此，研究蛋白质一级结构的差异，是研究生物进化的重要方法。蛋白质的一级结构，是基因遗传信息表达的产物，在进化过程中，随着基因突变，蛋白质一级结构中氨基酸残基的排列顺序也会发生变化，这在各种同源蛋白质一级结构的比较中清楚地显示出来。不同种属的生物中具有同一功能的蛋白质，在进化过程中可来自相同的祖先，但存在着种属差异，这是基因在进化过程中趋异突变（divergent mutation）的结果。

细胞色素 C（cytochrome C，cyt C）是研究蛋白质一级结构与分子进化关系及种属差异最好的实例。Cyt C 存在子细胞线粒体中，是生物氧化还原体系中的电子递体，它由单链蛋白质与血红素辅基组成。对 120 多个生物种属细胞色素 C 的一级结构分析发现，cyt C 一级结构中某些位置的氨基酸残基随物种的不同而有差异。物种间亲缘关系愈近则差异愈少，反之残基差异愈大。如人与黑猩猩的 cyt C 氨基酸残基顺序完全一致，而人与酵母的氨基酸残基则相差 44 个。因此，通过测定比较各物种蛋白质一级结构的差异程度，可以了解生物进化的基本规律。根据蛋白质中氨基酸残基的变化情况与古生物学测定生物进化年代相比较，可以发现各种蛋白质分子进化都有其本身的进化速率，称为单位进化周期（unit evolution period），也就是蛋白质在进化过程中每一个残基变异所需的时间。不同蛋白质的单位进化周期不同，细胞色素 C 的单位进化周期为 2640 万年，血红蛋白仅为 580 万年。

（三）蛋白质一级结构与分子病

蛋白质一级结构发生改变，甚至只有一个氨基酸残基发生改变，即可以使蛋白质的生理功能发生改变，严重时可导致疾病。这种由于基因突变而导致蛋白质一级结构改变后表现出生理功能异常，使机体出现病态现象，称为分子病。镰刀型红细胞贫血症是一种典型的，也是人们认识最早的一种分子病。1949 年 Pauling 发现镰刀型红细胞贫血症

是由于患者红细胞中含有一种异常血红蛋白（HbS）所致。1956 年 Ingram 采用指纹技术分析其一级结构，发现 HbS 与正常成人血红蛋白（HbA）的一级结构仅有一个氨基酸残基的差异，即在 β 链第 6 位上的谷氨酸被缬氨酸替代。血红蛋白 β_6 位的氨基酸位于 β 亚基的表面，当谷氨酸被缬氨酸取代后，亲水侧链变成疏水侧链，使血红蛋白的溶解度降低，在脱氧情况下易形成纤维状沉淀，导致与氧的亲和力降低，红细胞被扭曲成镰刀状，可塑性降低，细胞容易破裂，导致溶血性贫血。

第二节　蛋白质的空间结构与功能

蛋白质的空间结构又称为三维结构或空间构象，包括二级、三级和四级结构。随着蛋白质分子结构研究的深入，又提出超二级结构和结构域。蛋白质的空间结构是蛋白质特有性质和功能的结构基础。

一、蛋白质的二级结构

蛋白质的二级结构（secondary structure）是指蛋白质分子中肽链的局部空间结构，也就是多肽链骨架中氨基酸残基相对空间位置。蛋白质二级结构有 5 种基本的构象单元，①α－螺旋（α－helix）：多肽链骨架围绕中心轴顺时针方向螺旋式上升，每上升一圈为 3.6 个氨基酸残基，螺距为 0.54nm。根据主链骨架旋转方向不同，分为左手螺旋和右手螺旋；②β－折叠（β－pleated sheet）：呈折纸状，多肽链充分伸展，每个肽单元以 α－碳原子为旋转点，依次折叠成锯齿状结构，氨基酸残基侧链交替的位于锯齿状结构的上下方，均与片层相垂直。β－折叠有平行式和反平行式两类，前者肽链的 N－端都在同一端，后者所有肽链 N－端按正反方向排列，维持蛋白质空间结构的稳定性；③β－转角：常发生在肽链进行 180°回折时的转角上。肽链上一个残基的 C＝O 隔 2 个氨基酸残基与第四个残基的氨基氢形成氢键，维持 β－转角结构的稳定性；④无规则卷曲：指蛋白质分子中无规则卷曲的那部分肽链构象；⑤无序结构：指蛋白质分子中没有确定空间结构的区域。

蛋白质的二级结构以一级结构为基础，即一级结构中氨基酸残基的排列顺序决定二级结构的类型。此外，蛋白质空间构象的正确形成，还需一类称为分子伴侣（chaperon）的蛋白质参与。蛋白质初合成时，尚未折叠的肽段有许多疏水基团暴露在外，具有分子内和分子间聚集的倾向，使蛋白质不能形成正确空间构象。分子伴侣可逆地与未折叠肽段的疏水部分结合后松开，如此重复进行可防止错误的聚集发生，使肽链正确折叠。分子伴侣也可与错误聚集的肽段结合，使之解聚后，再诱导其正确折叠。

二、超二级结构和结构域

（一）超二级结构

蛋白质分子中两种主要的二级结构单元由于折叠盘曲，在空间上进一步聚集、组合在一起，形成有规则的二级结构聚合体，可作为结构域的组成单位，或直接作为二级结

构的“建筑块”，这种二级结构的聚合体称为超二级结构（super secondary structure）。常见的超二级结构有 αα、βββ、βαβ 三种类型（图 2－1）。αα 是相邻的 α 螺旋通过肽链连接而成，此种组合非常稳定，

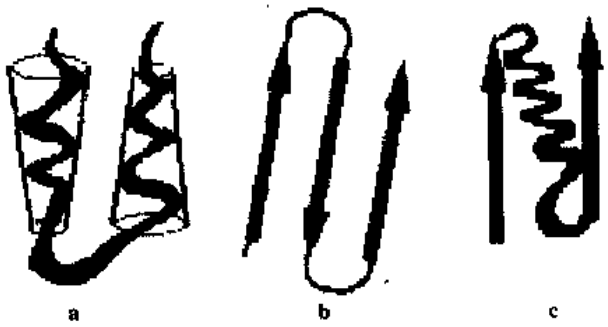


图 2－1 三种超二级结构示意图

a αα 结构 b βββ 结构 c βαβ 结构

常存在于 α－角蛋白和原肌球蛋白等纤维状蛋白之中；βββ 是由 3 条或 3 条以上的 β 折叠聚集而成，它们之间以短链相连，有时可由多条 β 折叠形成超二级结构；βαβ 是由一个 α 螺旋和与之首尾相邻的 β 折叠聚合而成，有时两组 βαβ 聚合在一起，形成更为复杂的超二级结构，它存在于许多球蛋白之中。

（二）结构域

在较大蛋白质分子中多肽链上相邻的超二级结构可形成紧密、稳定而且在蛋白分子构象上明显可分的区域，称为结构域（domain），各结构域之间的肽链松散、弯曲，形成分子内裂隙。结构域一般由 100～200 个氨基酸残基组成，各个结构域具有独特的空间结构，担负着不同的生理学功能。在多肽链折叠卷曲时，结构域又可能彼此靠近，形成球状蛋白分子。一般较大的蛋白质分子可含有两个或多个结构域，如 3－磷酸甘油醛脱氢酶的酶蛋白部分是由 2 个结构域组成，而 IgG 则含 12 个结构域。在 IgG 的 12 个结构域中，两条轻链上各有 2 个结构域，两条重链上各有 4 个结构域，补体结合部分与抗原结合部分处于不同的结构域。同一个蛋白质分子的结构域有的是相同的，也有的是不同的。不同蛋白质分子中可以有相似的结构域，如 3－磷酸甘油醛脱氢酶、乳酸脱氢酶、苹果酸脱氢酶等均属子以 NAD^+ 为辅酶的脱氢酶类，它们各由 2 个不同的结构域组成，其中与 NAD^+ 结合的结构域基本相同。由于结构域之间连接的柔韧性，彼此之间可有相对运动，这种运动对表达蛋白质的功能是很重要的。

三、蛋白质的三级结构

蛋白质的三级结构（tertiary structure）是指蛋白质分子中一条多肽链在二级结构、超二级结构和结构域的基础上进一步盘曲、折叠形成的空间结构，也就是整条肽链所有原子在三维空间的排布位置。三级结构靠次级键维系、二硫键加固。对单链蛋白质来

说，三级结构就是蛋白质分子的特征性立体结构，具有了蛋白质的特性和功能。而对多链蛋白质来说，具有三级结构的多肽链，称为蛋白质分子的亚基，尚不具有蛋白质的生物学活性。具有三级结构的蛋白质其分子或亚基一般呈似球状、椭球状或纤维状。其结构特点是整个分子比较致密，亲水性氨基酸残基大多分布在分子表面，形成亲水面；而疏水性氨基酸残基大多埋在分子内部，形成疏水核，少数在表面形成疏水面。在球状蛋白质分子表面常出现一个内陷的“洞穴”，该洞穴一般是疏水区，是蛋白质与配体结合发挥生物功能的中心部位。对酶而言，该洞穴则是与底物结合并发挥催化反应的酶的活性中心。

随着蛋白质三级结构研究的进展，在许多蛋白质分子中，可发现 2 个或 3 个具有二级结构的肽段，在空间上相互接近，形成立体形状或拓扑结构颇为类似的，具有特殊功能的局部区域，被称为模序（motif）。一个模序总有其特征性的氨基酸序列，并发挥特殊的功能。如许多钙结合蛋白质分子中通常有一个结合钙离子的模序，他由 α 螺旋－环－α 螺旋三个肽段组成（图 2－2），在环中有几个恒定的亲水侧链，侧链末端的氧原子通过氢键而结合钙离子。

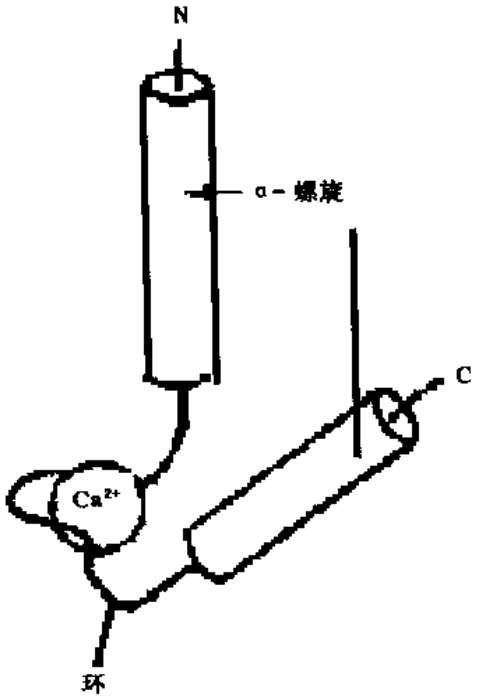


图 2－2　钙结合蛋白中的结合钙离子的模序

近年来发现的锌指结构（zinc finger）也是一种常见的模序例子。锌指结构由其形似手指，具有结合锌离子的功能而得名。该模序有一个 α 螺旋和 2 个反平行的 β－折叠三个肽段组成（图 2－3）。其 N－端有一对半胱氨酸（Cys）残基，C 端有一对组氨酸（His）残基，此 4 个残基在空间上形成一个洞穴，恰好容纳 1 个 Zn^{2+}。Zn^{2+} 分别与 2 个 Cys 和 2 个 His 以配位键相连，可稳定模序中的 α－螺旋结构，致使此 α－螺旋能镶嵌子 DNA 的大沟中，因此含锌指结构的蛋白质能与 DNA 或 RNA 结合，参与 DNA 的复制和转录。

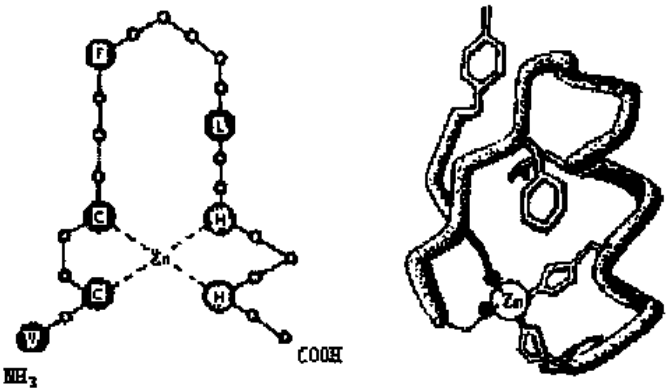


图 2—3 锌指结构示意图

四、蛋白质的四级结构

有些蛋白质是由一条多肽链盘曲折叠而成，此类蛋白质仅涉及一、二、三级结构。还有许多具有生物活性的蛋白质分子由两条或多条多肽链组成，每一条多肽链均具有其完整的三级结构，彼此以非共价键相连。此类蛋白质分子中的每条多肽链称为亚基（subunit），亚基与亚基之间呈特定的三维空间排布。这种蛋白质分子中各个亚基的空间排布和相互间的布局称为蛋白质的四级结构（quaternary structure）。蛋白质的四级结构不包括亚基内部的空间结构。在具有四级结构的蛋白质分子中，单独的亚基一般无生物学功能，只有具有完整的四级结构寡聚体才有生物学活性。在四级结构中亚基可以是相同的，也可以是不同的，前者称为均一四级结构，后者称非均一四级结构。血红蛋白是由 2 个 α 亚基和 2 个 β 亚基组成，两种亚基的三级结构相似，且每个亚基都结合有一个血红素辅基，各亚基凹凸镶嵌，组成一个球状分子（图 2—4）。四聚体的血红蛋白分子具有运输氧气和 CO_2 的功能，但每一个亚基单独存在时，虽可结合氧且与氧的亲和力增强，但在体内组织中难于释放氧。

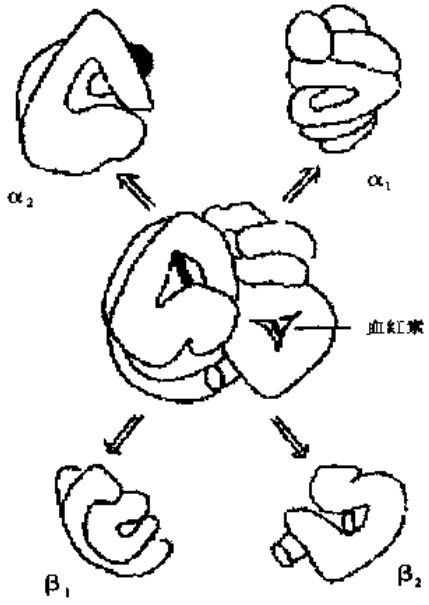


图 2—4 血红蛋白的三级及四级结构

五、蛋白质空间结构与功能的关系

蛋白质空间结构是蛋白质功能的基础。若蛋白质严密有序的空间结构遭到破坏，其生理结构也就丧失。现以酶催化作用以及血红蛋白携带氧的分子机制为例，阐述蛋白质空间结构与功能的关系。

（一）酶的空间结构与催化功能的关系

酶是具有催化作用的高分子化合物，其化学本质是蛋白质，同样具有一级、二级和三级结构，有些酶还具有四级结构。自然界存在的酶有数千种，各种酶具有特异的催化功能，这是由于各种酶的空间结构，特别是酶活性中心的空间构象不同所决定的。酶的活性中心是酶分子中具有三维结构的区域，或为裂缝或为凹陷。此裂缝或凹陷由酶的特定空间构象所维持，深入到酶分子内部，且多为氨基酸残基的疏水基团组成的疏水环境，形成疏水“口袋”。酶活性中心的空间构象是酶表现活力的关键部位，不同酶活性中心的空间结构不同，因而具有特异的催化活性。酶活性中心空间构象的维持则依赖于酶蛋白的二、三级结构的完整性。如果酶受各种变性因素，如高温、高压、强酸、强碱、重金属等的作用使空间结构破坏后，活性中心的构象也随着发生改变，酶也因之失活。

酶原与酶原激活是酶的空间结构与功能关系的最好证据。酶原之所以没有活性，是因为其尚不具有活性中心，当其受到激活因素的作用，蛋白质的三维空间结构发生改变，活性中心形或，也就获得了催化功能（见第五章）。酶变构调节的机制也是蛋白质空间结构与功能关系的例证。有些酶蛋白的构象在两种构象之间转换，一种构象是有活性状态，另一种是无活性状态。例如天冬氨酸转氨甲酰酶是由 2 个催化亚基三聚体和 3 个调节亚基二聚体（共 12 个亚基）组成。调节亚基可与 CTP 和 ATP 结合。该酶是体内合成嘧啶核苷酸的第一个关键酶，催化氨基甲酰基转到天冬氨酸上形成氨甲酰天冬氨酸。当 CTP 增多时，CTP 与调节亚基结合，可使酶成为无活性状态的构象，反之，若 CTP 减少，ATP 则可与调节亚基结合，使酶转变成有活性的状态。

（二）血红蛋白的构象变化与结合氧

血红蛋白(hemoglubin，Hb)是由 2 个 α 亚基和 2 个 β 亚基组成的四聚体，每一个亚基都结合一个血红素辅基。Hb 亚基之间通过 8 对盐键(图 2－5)使 4 个亚基紧密结合而形成亲水的球状蛋白。Hb 的功能是运输 O_2 和 CO_2，一分子 Hb 能结合 4 个氧分子，血红素铁就是氧结合部位。当 Hb 中第一个亚基与 O_2 结合以后，能促进第二及第三个亚基与 O_2 的结合，当前三个亚基均与 O_2 结合后，又大大促进第四个亚基与 O_2 结合，这种效应称为正协同效应(positi－ve cooperativity)。

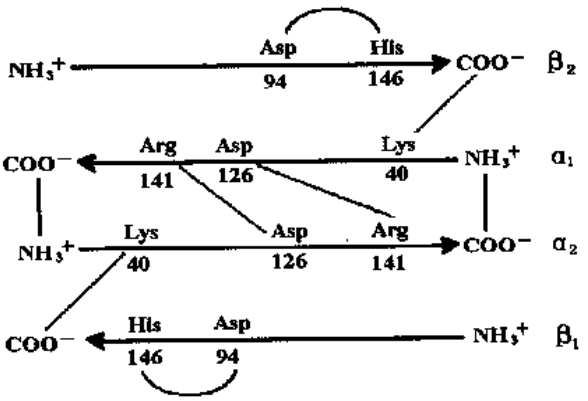


图 2－5　脱氧 Hb 亚基之间和亚基内的盐键

X 线衍射技术分析 Hb 和氧合 Hb 结晶的三维结构图谱发现 Hb 有两种状态。未结合 O_2 时，Hb 的 α/β 和 α/β 呈对角排列，结构较为紧密，与 O_2 的亲和力小，称为紧张态（tense state，T 态）。随着 O_2 的结合，4 个亚基羧基末端之间的盐键断裂，其二级、三级和四级结构也发生变化，使 α/β 和 α/β 的长轴形成 15°的夹角，结构显得相对松弛，

称为松弛态（relaxed state，R 态）。T 态转变成 R 态是逐个结合 O_2 而完成的。在脱氧 Hb 中，Fe^{2+} 半径比卟啉环中间的孔大，因此 Fe^{2+} 高出卟啉环平面 0.075nm，而靠近 F8 位组氨酸残基。当第一个 O_2 与血红素 Fe^{2+} 结合后，使 Fe^{2+} 的半径变小，进入到卟啉环中间的小孔中，引起 F 肽段等一系列微小的移动，同时影响附近肽段的构象，造成 2 个 α 亚基间盐键断裂，使亚基间结合松弛，可促进第二个亚基与 O_2 结合，依此方式可影响第三、四个亚基与 O_2 的结合，最后使四个亚基全处于 R 态。此种一个氧分子与 Hb 亚基结合后引起的构象变化，称为变构效应（allosteric effect）。小分子 O_2 称为变构剂或效应物，Hb 则被称为变构蛋白。变构效应不仅发生在 Hb 与 O_2 之间，一些酶与变构剂的结合、配体与受体的结合也存在着变构效应，所以它具有普遍意义。

六、蛋白质—蛋白质相互作用

蛋白质—蛋白质相互作用（protein—protein interaction）是指蛋白质通过分子之间的相互作用实现其特异生物学功能。例如，酶与底物、抗原与抗体、配体与受体之间的相互作用等等。

（一）蛋白质分子之间的相互聚合

有些蛋白质分子只有与其他蛋白质相互作用或相互聚合才能发挥生物学作用。例如，抗原与抗体结合产生免疫反应；蛋白类激素与相应受体结合发挥生物学效应；相关功能的酶分子聚合在一起，形成具有一定功能的多酶复合体，如丙酮酸脱氢酶复合体由三种酶聚合而成。

一个蛋白质分子与配体之间的结合需要在两个分子的特定部位之间形成多个非共价键，两分子间相互作用的部位称为结合位点（binding site）。结合位点通常是由多肽链上一些相互分离的氨基酸残基通过肽链折叠而在空间上聚集到一起，形成特定的空间排列方式。这些氨基酸残基仅仅属于多肽链的一小部分，其余的氨基酸残基为维持其正确空间位置所必需，并提供调节过程所需要的其他结合位点；蛋白质内部氨基酸残基则对维持蛋白质分子适当形状和刚性起作用。

（二）蛋白质相互作用的特异性

蛋白质分子之间的相互作用具有特异性，即一种受体蛋白只能与一种特异配体蛋白结合，这种特异性是由其结合位点决定的。两个分子的结合位点必须存在能够配对形成非共价键的基团，而且每对基团在空间上恰好能达到形成非共价键的最佳距离。这就是所谓的“镜像关系”，即两分子的结合位点具有相对应的基团和对称的空间排列。这种基团的对应及空间排列对称关系决定了蛋白质分子之间相互作用的特异性，从而保证了生物分子能正常发挥其生物学功能。如信号传递过程中，信号分子（如激素或神经递质等）与特异受体蛋白的特定位点结合后，导致受体分子构象变化，启动信号的传递，引起细胞功能变化。细胞或组织对特异配体分子的应答是由其所具有的特异受体及其与配体结合所启动的胞内反应完成的。

（高　聆）

第三章　基因重组技术

不同来源的 DNA 分子通过末端共价连接形成重新组合的 DNA 分子的过程称为 DNA 重组（DNA recombination）。DNA 重组是生物细胞内广泛存在的自然现象，是生物遗传和进化的分子基础。自从 1973 年 Stanley Cohen 等人首次采用质粒与外源 DNA 体外重组的方法克隆 DNA 以来，重组 DNA 技术（DNA recombination technique）作为分子生物学的一项重要技术得到了迅速发展，使科学家们分离、分析和操作基因的能力几乎达到无所不能的地步。

重组 DNA 技术又称基因重组技术或基因克隆（gene cloning）技术。所谓克隆（Clone）是指通过无性繁殖过程产生的在基因型和表型上与亲代完全相同的子代群体。基因克隆则是指在体外对 DNA 分子按照既定的目的分离、剪切和重新连接，构成重组的 DNA 分子，然后把它导入宿主细胞，使其在宿主细胞内扩增，从而获得该 DNA 分子的大量拷贝。由于此克隆技术是在分子水平上进行的，故又称为分子克隆（molecular cloning）。基因克隆技术可用于定点改造蛋白质分子中氨基酸的排列顺序，从而改变蛋白质的理化性质和生物学活性，因此，基因克隆技术是蛋白质工程中常用技术。

第一节　工具酶

DNA 分子的剪切、重组、合成及修饰涉及一系列酶促反应，催化这些反应的酶是基因操作的基本工具，故又称为工具酶。工具酶在基因克隆中占有非常重要的地位，只有了解其作用机制后，才能加以灵活应用。

一、限制性核酸内切酶

限制性核酸内切酶（restriction endonucleases）又称限制性内切酶，简称为限制酶。这类酶能识别双链 DNA 分子内部的特异位点并且裂解磷酸二酯键。

（一）限制酶的命名和分类

1. 命名　限制酶的命名采用 Smith 和 Nathane 提议的命名系统。第一个字母取自产生该酶的细菌属名，用大写；第二、第三个字母是细菌的种名，用小写；第四个字母代表株名。另外用罗马数字代表同一菌株中不同限制酶的编号，现在常用来表示发现的

先后顺序。例如 HindⅢ H 代表 *Haemophilus* 属，*in* 代表 *influenzae* 种，d 代表 Rd 菌株，Ⅲ代表该菌株中第三个被分离到的限制酶。

2. 分类 目前发现的限制酶有数百种之多，根据酶的结构、识别切割序列上的特异性、催化条件及修饰活性等，将限制酶分为Ⅰ，Ⅱ，Ⅲ型。Ⅰ型酶为多亚基蛋白复合物，兼具限制酶和甲基化酶的活性，辅助因子除 Mg^{2+} 外，还需要 ATP 和 S－腺苷甲硫氨酸（SAM）的参与。Ⅰ型酶通常在距识别位点下游 100～1000bp 处随机切割 DNA 分子。因其切割作用缺乏特异性，故在 DNA 重组工作中并没有什么实际的用处。现已知的Ⅲ型酶数量相当少，由两个亚基组成，与Ⅰ型酶类似，也兼具限制酶和甲基化酶活性，催化活性中也需要 Mg^{2+}，ATP 和 SAM 为辅助因子。虽然Ⅲ型酶能在识别位点附近切割 DNA，但其切割位点很难确定，故在 DNA 重组工作中也没有实际用处。Ⅱ型酶由两个相同的亚基组成，仅具有限制作用，反应的辅助因子为 Mg^{2+}，对 DNA 的识别和切割具有极强的特异性，而且其切割的序列可知、固定，故是 DNA 重组技术中最重要的工具酶。通常所指的限制酶就是Ⅱ型酶。

（二）Ⅱ型限制酶的识别和切割位点

Ⅱ型酶的识别序列通常为 4～6bp，具有回文结构（palindrom），大多数酶是错位切割双链 DNA，产生 5′或 3′末端突出的黏性末端（sticky end）。

如 EcoRI 切割后产生 5′末端突出的黏性末端：

5′－G↓AATT C－3′　　5′－G　　AATTC－3′

3′－C TTAA↑G－5′　⟶　3′－CTTAA　　G－5′

PstI 切割后产生 3′末端突出的黏性末端：

5′－C TGCA↓G－3′　　5′－CTGCA　　G－3′

3′－G↑ACGT C－5′　⟶　3′－G　　ACGTC－5′

还有一些酶沿对称轴切割 DNA，产生平端或钝端（blunt end）。如 SmaI：

5′－CCC↓GGG－3′　　5′－CCC　　GGG－3′

3′－GGG↑CCC－5′　⟶　3′－GGG　　CCC－5′

不同限制酶识别 DNA 中核苷酸序列长短不一，有的识别序列是 4 个核苷酸，有的是 6 个或 8 个核苷酸。识别序列的大小决定其切割后产生 DNA 片段的长短。如果 DNA 序列是随机的，识别序列为 4bp 的限制酶其切口几率应为 $(1/4)^4=1/256$，即 DNA 链上每当 256 个 bp 的长度可能含有一个切口；如果识别序列为 6bp，则切口几率为 $(1/4)^6=1/4096$。这种可能性随 GC 含量变化而变化。不同酶切割 DNA 频率不同，切割后产生黏性末端长短不一样，性质也不同，这对重组 DNA 或有关分子生物学操作及应用影响很大。

有些限制酶识别序列不同，但切割 DNA 后产生相同类型的黏性末端，这些酶称为同尾酶（isoaudamers），如 BamHI（G↓GATCC）和 Sau3AI（N↓GATCN）。由此产生的黏性末端称为配伍末端（compatible end），可进行相互连接，在 DNA 重组中具有更大灵活性。

有些限制酶来源不同，但能识别和切割同一位点，这些酶称为同功异源酶（isos-

chizomers)。如 Bam HI 和 Bst I 识别和切割位点均是 G↓GATCC，XhoI 和 Pac R7 则同为 C↓TCGAG，这些同功异源酶可以互相代用。

现已提纯供商品用的限制酶有 400 余种，其中常用者列举于表 3－1。使用限制酶时应注意提供适宜的反应条件，如 DNA 底物的纯度、浓度、分子结构和构型、缓冲液的 pH 及离子强度等。如果反应体系不能满足最适条件，则可能导致酶产生第二活性，又称星活性（star activity），即限制酶严格的识别特异性降低，导致其在 DNA 分子内产生附加切割。如 EcoRI，当反应条件改变时，其特异性可由原来 6bp 的 GAATTC 降为 4bp 的 AATT，一般用 EcoRI * 表示，称为 EcoRI 的星活性。

表 3－1　常用的限制酶

酶	辨认的序列和切口	说明
AluI	……AG↓CT…… ……TC↑GA……	四核苷酸，平端切口
SmaI	……CCC↓GGG…… ……GGG↑CCC……	六核苷酸，平端切口
BamHI	……G↓GATCC…… ……CCTAG↑G……	六核苷酸，粘端切口
BglI	……A↓GATCT…… ……TCTAG↑A……	六核苷酸，粘端切口
EcoRI	……G↓AATTC…… ……CTTAA↑G……	六核苷酸，粘端切口
HindIII	……A↓AGCTT…… ……TTCGA↑A……	六核苷酸，粘端切口
SalI	……G↓TCGAC…… ……CAGCT↑G……	六核苷酸，粘端切口

二、其他工具酶

（一）DNA 连接酶

DNA 连接酶（DNA ligase）的作用是催化 DNA 片段之间的 5′磷酸基和 3′羟基间形成 3′，5′—磷酸二酯键，主要用于连接两个独立的 DNA 片段或修复双链 DNA 中一条链上的切口。现常用的有 T4 DNA 连接酶和大肠杆菌 DNA 连接酶。T4 DNA 连接酶最早是从 T4 噬菌体感染的大肠杆菌中发现并提取的，是一种既可以连接黏性末端，又

可以连接平头末端的 DNA 连接酶，催化反应需要 ATP 作为辅助因子并提供能量。大肠杆菌 DNA 连接酶只能连接黏性末端，辅因子为 NAD^+。

（二）DNA 聚合酶

基因克隆中常用的 DNA 聚合酶（DNA polymerase）有大肠杆菌 DNA 聚合酶Ⅰ、Klenow 片段和 T4 DNA 聚合酶。

大肠杆菌 DNA 聚合酶Ⅰ是一种多功能酶，包括三种活性：① 5′→3′聚合酶活性：以 DNA 为模板，以四种脱氧核苷酸（dATP、dGTP、dCTP 和 dTTP）为底物，在 Mg^{2+} 参与下，在引物或 DNA 缺口的 3′－OH 上合成 DNA，使 DNA 链沿 5′→3′方向延长；② 5′→3′外切酶活性：从游离的 5′－末端降解双链 DNA 成为单核苷酸；③ 5′→3′外切酶活性：从游离的 3′－末端降解双链或单链 DNA 成为单核苷酸，但对双链 DNA 3′→5′外切活性可被 5′→3′聚合活性所抑制。由于它具有 5′→3′核酸外切酶活性，当用缺口平移法（nick translation）标记 DNA 探针时，常用 DNA 聚合酶Ⅰ。

用枯草杆菌蛋白酶可将 DNA 聚合酶Ⅰ裂解成分子量为 68kD 和 36kD 的两个片段。其 C 端的大片段称为 klenow 片段（Klenow fragment），具有 5′→3′聚合酶活性和 3′→5′外切酶活性。其 3′→5′外切酶活性能保证 DNA 复制的准确性，把 DNA 合成过程中错误配对的核苷酸去除，再把正确的核苷酸接上去。Klenow 片段的主要用途有：①补齐双链 DNA 的 3′－末端；②通过补齐 3′－端使 3′－末端标记；③在 cDNA 克隆中合成第二股链；④ DNA 序列分析。T4 DNA 聚合酶的活性与 Klenow 片段类似，但其 3′→5′外切酶活性较大。

（三）末端脱氧核苷酰转移酶

末端脱氧核苷酰转移酶（terminal deoxy nucleotidyl transferase，TdT）简称末端转移酶，它的作用是将脱氧核苷酸逐一加到 DNA 的 3′－OH 上。主要用于在载体和待克隆的片段上形成同聚物尾，构建便于重组的人工黏性末端，以便于进行连接。其次也可以用于 DNA 片段的末端标记。

第二节 载 体

外源 DNA 本身没有自主复制能力，导入宿主细胞后不能随宿主细胞的繁殖而复制，达不到使外源 DNA 片段扩增的目的。如果要将外源 DNA 导入宿主细胞进行扩增成表达，就需要将外源 DNA 与某种能在宿主细胞中进行自主复制和表达的 DNA 重组，并由其携带进入宿主细胞。这种可以用来插入外源 DNA 片段构建重组 DNA 分子，并将外源 DNA 携带进入宿主细胞的 DNA 称为基因克隆载体，简称为载体（vector）。

一、质　粒

（一）质粒 DNA 生物学特性

质粒（plasmid）是一些染色体外的环形双链 DNA 分子。提纯后的质粒 DNA 分子有三种存在形式：①共价闭环 DNA（covalently closed circular DNA，ccDNA），以超螺旋形式存在；②开环 DNA（open circular DNA，oc DNA），此种质粒 DNA 的两条链中，有一条链保持着完整的环形结构，另一条链发生一处或多处断裂，形成缺口，可以自由旋转，从而消除了张力，形成松弛的环状分子；③线状 DNA（linear DNA），因质粒 DNA 的两条链在同一处断裂而形成，线状 DNA 又称为 L 构型（图 3－1）。电泳时，同一质粒的 DNA，其超螺旋状态的泳动速度最快。

质粒主要发现于细菌、放线菌和真菌细胞中，大小从 1kb 至 200kb 不等，具有自主复制能力，能在子代细胞中保持恒定的拷贝数，并表达所携带的遗传信息。质粒 DNA 复制和转录要依赖于宿主细胞编码的某些酶和蛋白质，因此，是一种寄生性的自主复制子。由质粒产生的表型包括对抗生素的抗性；产生某些抗生素；降解某些有机物以及产生某些酶等。

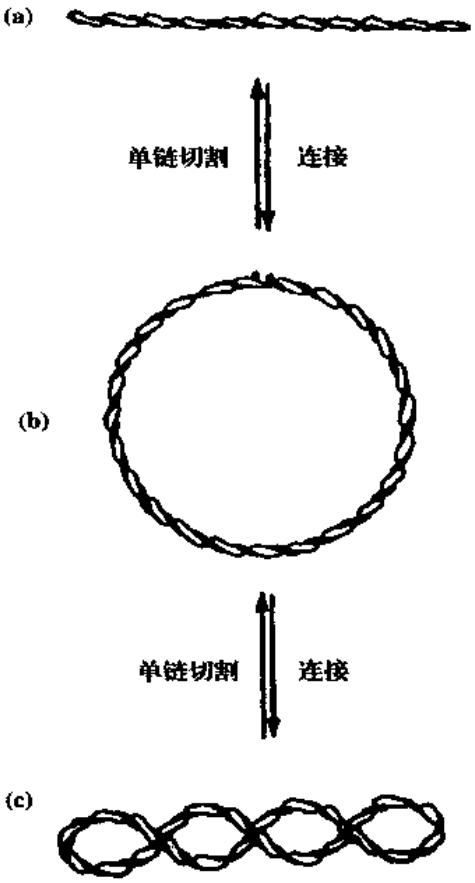


图 3－1　质粒 DNA 的分子构型
(a) 松弛线型的 L 构型
(b) 松弛开环的 OC 构型
(c) 超螺旋的 SC 构型

（二）良好质粒载体应具备的条件

质粒有许多类型，作为理想的克隆载体必须具备下列条件：①分子量相对较小，能在细菌内稳定存在，有较多的拷贝数；②具有可供选择的遗传标志，便于对宿主细胞进行识别和筛选，如抗生素的抗性基因、编码 β－半乳糖苷酶的 lac Z 基因等；③带有尽可能多的限制酶单一酶切位点，即多克隆位点（multiple cloning sites，MCS）。便于有选择性的外源基因定点插入。如果这些位点有外源基因的插入，会导致某些标志基因的失活，而便于筛选；④具有复制起始点（origin，ori），能在宿主细胞中自主复制，并能携带外源 DNA 片段一同扩增。

（三）常用的质粒载体

目前，经过人工构建，已有一系列符合良好载体条件的质粒作为商品供应，其中被广泛应用于基因克隆的质粒载体有 pBR322 和

pUC 系列。

1. pBR322 质粒载体　早期应用的质粒载体是由 Boyer 及其同事构建，当时他们构建起一套 pBR 质粒系列，pBR322 则是其中的一种。pBR322 质粒载体的特点是：①分子量较小，长度为 4.3kb，不仅易于自身 DNA 的纯化，而且能克隆较大外源 DNA，也不影响其复制；②具有氨苄青霉素和四环素两种抗生素的抗性基因，可供作转化子的选择标志。在两种抗生素抗性基因中间存在限制酶切位点，便于外源基因的插入和筛选（见图 3－2）；③具有较高的拷贝数，而且经过氯霉素扩增后，每个细胞中可积累 1400～3000 个拷贝。这为重组体 DNA 的制备提供了极大方便。

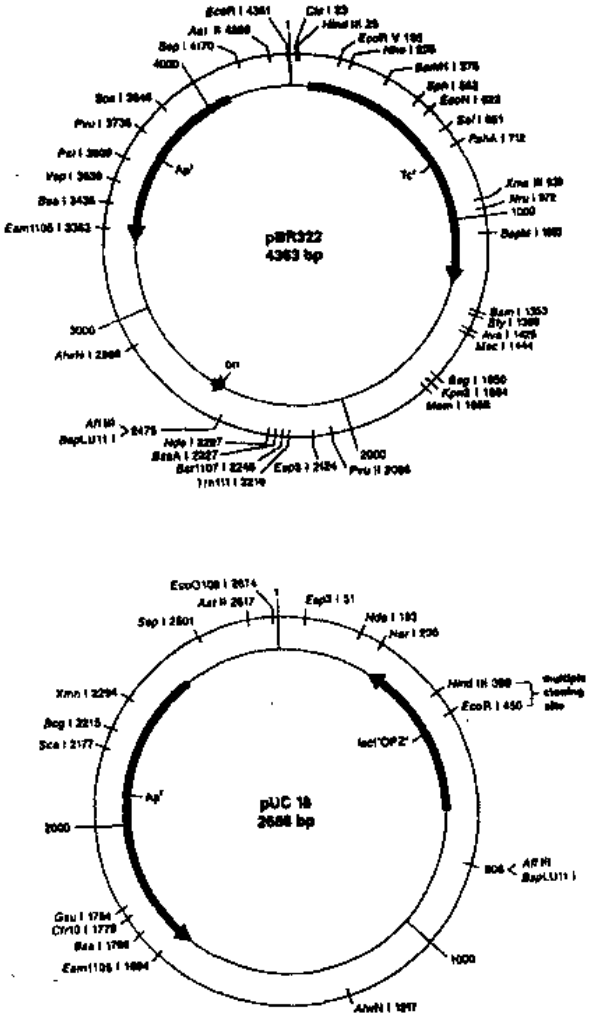


图 3－2　pBR322 质粒和 pUC18 质粒的物理图谱

2. pUC 质粒载体　pUC 质粒载体是在 pBR322 质粒载体的基础上，组入了一个在其 5′一端带有一段多克隆位点的 lac Z′基因，而发展成为具有双功能检测特性的新型质粒载体系列。pUC 质粒载体是由 J. Messing 和 J. Vieria 于 1987 年首先构建，包括四个组成部分（见图 3－2）：①来自于 pBR322 质粒的复制起始点（ori）；②氨苄青霉素抗性基因（Amp^r），其 DNA 核苷酸序列不再含有原来的限制性核酸内切酶的单识别位点；③大肠杆菌 β－半乳糖苷酶基因（lacZ）的启动子及其编码 α－肽链的 DNA 序列，此结构特称为 lacZ′基因；④位于 lacZ′基因中的靠近 5′一端的一段多克隆位点（MCS）区段，但它并不破坏该基因的功能。

pUC 质粒载体的优点是：①具有更小的分子量和更高的拷贝数，如 pUC18 的长度仅为 2.69bp，pUC18 质粒不经氯霉素扩增，平均每个细胞即可达 500～700 个拷贝；②因 lacZ′基因所编码的 α－肽链参与 α－互补作用，因此，适用于组织化学方法检测重组体；③在 lacZ 基因中加入了多克隆位点 MCS 区段，这使得具有两种不同黏性末端的外源 DNA 片段，无须借助其他操作而直接克隆到 pUC 质粒载体上。

二、噬菌体载体

（一）λ噬菌体载体

1. λ噬菌体（λphage）的生物学特性　λ噬菌体是一种大肠杆菌噬菌体，由头、尾两部分组成。头部呈六角型，装载了噬菌体的整个基因组（即λDNA）。λDNA 为双链线状分子，长约 49kb，在分子两端各有 12bp 的互补单链（是天然的黏性末端），称 cos 位点。当噬菌体吸附于宿主细胞膜后，λDNA 自噬菌体的尾部注入宿主细胞，其两个黏性末端互相结合，既可以裂解性生长，也可以溶解性生长。

λ噬菌体含有 60 多个基因，大多数基因的序列已经确定。其基因组可划分为三个区（见图 3－3）：①左侧区自基因 A 到基因 J，包括参与噬菌体头部蛋白质和尾部蛋白质合成所需要的全部基因；②中间区介于基因 J 与基因 N 之间，这个区段包含的基因对于噬菌体的生长关系不大，它们可以缺失或被代替而对噬菌体感染周期没有明显影响，故称为非必需区。但该区包括了一些与重组有关的基因，（例如 redA 和 redB），以及噬菌体整合到大肠杆菌染色体中的 int 基因和把原噬菌体从寄主染色体上删除下来的 xis 基因；③右侧区，位于基因的右侧，包括全部主要的调控成分，噬菌体的复制基因（O 和 P）以及溶菌基因（S 和 R）。

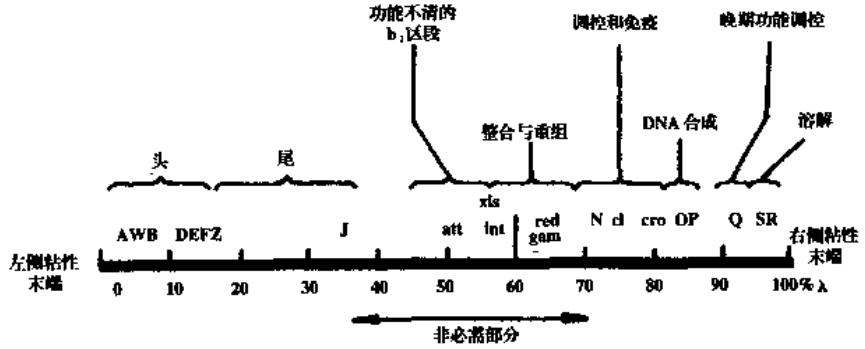


图 3－3　λ噬菌体 DNA 的结构

λDNA 必须包装上蛋白质外壳后才能感染大肠杆菌，只有位于野生型基因组长度 75%～105%范围的 λDNA 才能被包装成噬菌体颗粒。成熟的噬菌体粒子使细胞逐渐裂解，释放出许多新的具有感染性的病毒粒子。λ噬菌体在宿主细胞的另一种生活周期是溶源化。cI 产物阻遏 o，p 基因，而 int 表达导致 attB 和 attP 间的位点专一重组。随后便和染色体 DNA 一样进行复制，并随着细胞分裂转移到子代细菌中去。

（二）λ噬菌体载体

野生型噬菌体经突变和缺失等手段改造，已衍生出上百种克隆载体。目前常用的 λ 衍生物载体有插入型载体和置换型载体。

1. 插入型载体　将野生型 λ 噬菌体的 λDNA 中多余限制酶切位点删除，使其中心仅留有几个单一酶切位点。这样经限制酶切割的 λDNA 便能够与经同样限制酶消化的

外源 DNA 片段结合，使之插入在这个位点上，而不导致噬菌体功能的丧失。插入型载体装载容量较小，只能插入比较小的外源 DNA，如 λgt11 载体（见图 3－4）仅能克隆 7kb 以下的外源 DNA。但由于 λgtⅡ载体含有 lacZ 基因，插入位点 EcoR Ⅰ在其编码区的 3′端，当插入的 cDNA 阅读框与 lac Z 基因相一致时，能产生融合蛋白质，可以用免疫学方法，如蛋白质印迹分析进行检测。此外，重组的 λgt11 因 lac Z 基因失活，故在含 X－gal 培养基中形成白斑，而未重组的 λgt11 则形成蓝色斑点，便于区分筛选。λgt11 载体适用于建立 cDNA 文库。

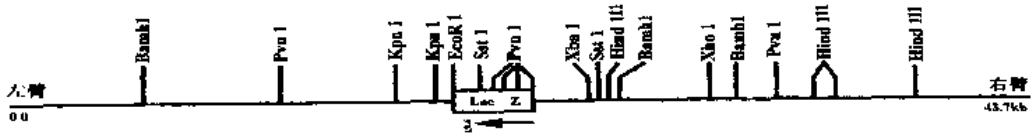


图 3－4 λgt11 载体的简图

2. 置换型载体 λDNA 中心区段是一些非必需序列，将该区段删除并替换上一个两侧多克隆位点是反向重复序列的外源 DNA 片段，即可建成置换型克隆载体。当外源 DNA 插入时，一对克隆位点之间的 DNA 片段即会被置换掉，从而有效地提高了克隆外源 DNA 片段的能力。如 EMBL4 中，长度为 13kb 的中间可取代片段有两个 Sal Ⅰ位点，包围其两侧的一对反向重复的多聚衔接物中存在着 EcoRI、BamHI 及 SalI 三种限制酶的单识别位点（见图 3－5），外源 DNA 的大小为 9～23kb。如果无外源 DNA 取代，由 λDNA 左右臂连接起来的 DNA 分子，由于太小不能被包装，这样就提供了一个挑选重组 λDNA 的阳性标志。

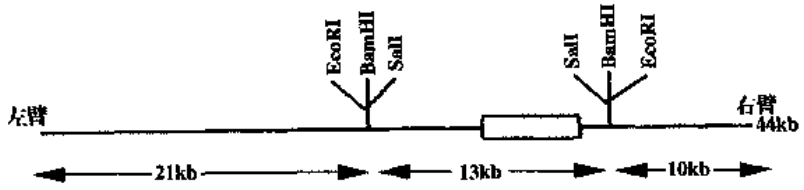


图 3－5 EMBL4 结构简图

（三）黏性质粒

λ 噬菌体载体适用于构建基因文库，但其装载容量不超过 20kb，对于更大的基因片段则无能为力。基于 DNA 被包装时的识别序列仅为 cos 附近很小区域的事实，J. Collins 和 B. Hohn 等人构建出黏性质粒载体，又称为柯斯质粒载体（cosmid vectors）。黏性质粒是由 λDNA 的 cos 区与质粒重新构建的载体，为双链环状 DNA，其克隆容量可达 40～50kb。

黏性质粒具有下列特点：①含有 λ 噬菌体的黏性末端 cos 区，所以它可以像 λ 噬菌体一样进行体外包装；②带有质粒的抗药性标记如 Tet^r 或 Amp^r 基因及自主复制成分，所以黏性质粒可以在细菌中大量复制扩增。当体外重组的黏性质粒分子包装成噬菌体颗粒并感染大肠杆菌后，可按质粒的方式进行复制；③具有一个或多个限制酶的酶切位

点；④其本身分子量较小，如 pHC79 仅 6.5kb，可容纳 40kb 左右的 DNA 片段；⑤由于非重组体黏性质粒很小，不能在体外包装，因而体外包装的主要是重组体，有利于以后的筛选。

（四）M13 噬菌体

M13 噬菌体（M13 phage）是一种大肠杆菌噬菌体，其基因组为 6.5kb 的闭合环状 DNA。感染型 M13 噬菌体为单链 DNA（single stranded DNA，ssDNA）（见图 3－6），当其感染大肠杆菌后，在菌体内借助宿主的酶系统先把 ssDNA 复制为双链 DNA（dsDNA），称为复制型 DNA（replicational form DNA，RF－DNA），可用作克隆载体。RF－M13 在菌体内继续繁殖至每个菌体达 100～200 个拷贝后，噬菌体内 DNA 的合成就变得不对称，有一种噬菌体蛋白附着在新合成的正链上，于是不能再合成新的双链 DNA，只有其中一条链进行复制，产生大量的单链 DNA，并被包装成噬菌体颗粒，然后从宿主菌中分泌出来，并再度感染其他宿主细胞。胞内的 RF 型分正负链，在胞外正负链均可作为模版进行半保留复制。感染型的单链以正链为主。

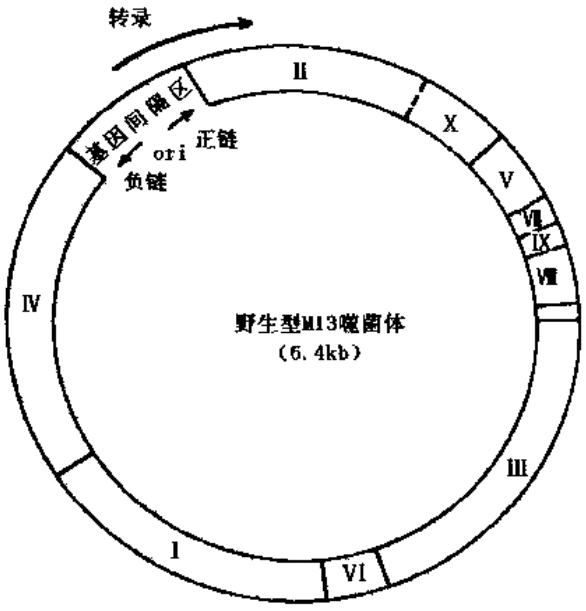


图 3－6　野生型 M13 单链 DNA 噬菌体的基因图

M13 噬菌体的最大优点是噬菌体颗粒中所含的是单链 DNA，可作为模版用于 DNA 序列分析。另外，利用单链 M13 克隆可制备成单链 DNA 探针，用于检测 DNA 或 RNA 的杂交分析；也可以用作基因定位诱变的载体。噬菌体成熟后分泌出细胞，可以从细胞培养液的上清液中获得噬菌体，制备单链噬菌体 DNA。双链噬菌体 DNA 可通过裂解细胞进行制备。

三、表达载体

表达载体（expressing vector）是用来在宿主细胞中表达外源基因的载体。这类载体除具有克隆载体所具备的性质外，还带有表达构件——转录和翻译所必需的 DNA 序列。表达载体包括原核细胞表达载体和哺乳动物细胞表达载体两类。

（一）大肠杆菌表达载体

一个良好的大肠杆菌表达载体（图 3－7）应具备克隆载体所具备的特性，如具有为载体提供表型选择标记的抗菌素抗性基因（Ter^r 或 Amp^r）决定载体拷贝数的复制起始点（ori）以及克隆位点等，还应含有表达系统元件即启动子、核糖体结合位点终止子等。

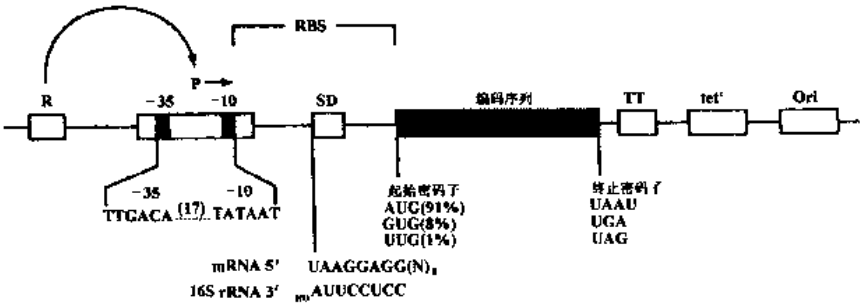


图 3－7 大肠杆菌表达载体的基本结构特征

1. 启动子（promoter） 大肠杆菌表达载体常用的启动子包括 trp－lac 启动子、λ噬菌体 PL 启动子和 T7 噬菌体启动子。

trp－lac 启动子，又称为 tac 启动子，是一个双启动子或称杂合启动子，由 trp 启动子加上 lac 操纵子中的操纵基因、SD 顺序融合而成。整个 tac 启动子受 lac 阻抑物调控，在 lac 阻抑物高水平表达的 lacIq 大肠杆菌菌株中，其转录可被抑制，异丙基硫代半乳糖苷（IPTG）可诱导其表达。

入噬菌体 PL 启动子是一种温度诱导的启动子，受控于温度敏感的阻抑物（cIts857）。（Its85）在低温下（37℃）可以阻抑 PL 启动子的转录，但在高温下（42℃）则失去阻抑作用。含 PL 启动子的表达载体需转化到 M5219 菌株中才能调控表达；常采用 30℃扩增，40℃诱导以减少热体克蛋白的产生。

T7 噬菌体启动子是一个表达效率很高的启动子，但需要特殊的受体菌。现在已有 JM109（DE3）等，是溶原菌，带有 lacUV5 启动子控制下的 T7 噬菌体 RNA 聚合酶基因，可用 IPTG 诱导。

2. 转译核糖体结合位点 mRNA 在细菌中的翻译效率依赖于是否有核糖体结合位点（ribosome--binding site，RBS）的存在。RBS 包括起始密码（ATG）和 SD 序列。SD 序列是由 Shine－Dalgarno 于 1974 年发现，富含嘌呤核苷酸，刚好与 16SrRNA3′－末端富含嘧啶的序列互补，mRNA 子是与核糖体很好结合。也有人将所构建的载体中的 SD 序列称为 RBS 位点，或 RBS 序列。RBS 是大肠杆菌表达载体中必不可少的元件。

3. 终止子 一个基因 3′－末端的能被 RNA 聚合酶识别并停止转录的 DNA 序列称为转录终止子。原核生物终止子序列有一段富含 A/T 的区域和一段富含 G/C 的区域，富含 G/C 区域具有回文对称结构。这种终止子转录后形成的 RNA 具有茎环结构，当它与 RNA 聚合酶作用时，导致酶的构象改变，使酶不再向下游移动，终止 RNA 的合成。

（二）哺乳动物表达载体

克隆基因要在哺乳动物细胞中进行表达，必须将基因重组到适当的真核表达载体中。哺乳动物表达载体含有必不可少的原核序列，如在大肠杆菌中能起作用的复制起始位点以及便于筛选重组质粒的抗生素抗性基因等，还含有在真核细胞中工作的遗传元件，如增强子、启动子、转录终止序列以及加多聚腺苷酸的信号序列等。

1. 增强子—启动子　一般来说原核启动子在哺乳动物细胞中是不起作用的，但有些来源于动物病毒的启动子和增强子能被真核细胞识别。他们不但可以引发动物基因工程中常用的一些标记基因的表达，而且能够引发克隆的外源基因的表达。如 SV40 病毒早期基因增强子（SV40）、Rouse 肉瘤病毒基因组长末端重复序列（RSV）、人类原细胞病毒（heMV）等。这些增强子—启动子组合可在广泛的宿主细胞中作用，是目前真核表达载体中常用的，如 hcMV 的增强子—启动子已被用于构建成哺乳动物表达载体 pC MV－Script。

2. 转录终止和加 poly（A）信号　真核基因表达的过程中，RNA 聚合酶Ⅱ通常跨过 poly（A）位点继续进行转录，因此，成熟的 mRNA 3′－端是经过位点特异性的转录后切割并加上 poly（A）而形成的。

准确而有效地加上 poly（A），有赖于 mRNA 中不同的序列，一是位于 poly（A）位点下游的 GU 富集区，二是 poly（A）加尾信号：AAUAAA。真核表达载体必须带有 poly（A）位点下游序列，以保证新转录的 mRNA 能够有效地加上 poly（A）。尽管全长 cDNA 克隆可能已带有 AATAAA 序列和一段 poly（A），但这些内源性序列本身并不足以保证 poly（A）的形成。因此，载体中必须包含切割和加 poly（A）所必不可少的下游 GU 富含区。最为常用的加 poly（A）信号来自 SV40，是一段 237bp 的 BamHI－BclI 限制酶切片段，其中同时含有早期和晚期转录单位的切割与加 poly（A）信号，两套信号作用的方向相反，并分别位于不同的 DNA 链上，对 mRNA 的加工均很有效。

第三节　基因克隆的基本程序

基因克隆的基本程序包括五大步骤：①制备目的基因；②连接目的基因与载体形成重组体（重组子）；③将重组体转入宿主细胞；④重组体的筛选和鉴定；⑤重组体的扩增和表达.

一、目的基因的制备

目的基因是指要克隆或表达的基因，又称外源基因。目的基因可以有如下几种来源和制备方法：

（一）从基因组文库中制备

采用限制酶将基因组 DNA 切成片段，与载体连接成重组体，重组 DNA 导入适当细胞中扩增，获得各 DNA 片段的克隆，这一组克隆混合体称为基因组或基因文库（genomic library，g 文库）。基因组文库包括某种细胞器或某个染色体基因组所有 DNA 片段，即全部遗传信息。目的基因可以从基因组文库中筛选。

（二）从 cDNA 文库中制备

从生物细胞中提取总 RNA，通过反转录酶反应合成与 mRNA 互补的 DNA，即为

cDNA，将 cDNA 混合体与载体连接形成重组 DNA 混合体，重组 DNA 导入宿主细胞进行扩增，得到分子克隆的混合体，该混合体称 cDNA 文库（c 文库）。cDNA 文库与 gDNA 文库不同，它只有表达的转录体互补序列，不包括内含子和其他表达的 DNA 序列。从 cDNA 文库中也可以筛选出所需的目的基因。

（三）PCR 扩增目的基因

采用聚合酶链反应（PCR）技术（见第四章）可以在体外特异地扩增目的基因片段。在已知序列前提下，也可以采用反转录 PCR（RT－PCR）方法直接从 mRNA 获得特定基因的 cDNA。PCR 扩增目的基因的方法，较从基因组文库或 cDNA 文库获得目的基因更为简便、快捷和准确。

（四）人工合成目的基因

如果知道肽链的氨基酸顺序，则可按照对应的密码子推导出 DNA 的碱基序列，然后用化学方法人工合成这段序列。DNA 合成仪的诞生，为人工合成目的基因提供了方便、快捷的合成方法。

二、目的基因与载体重组

目的基因片段与适当的载体经限制酶剪切后，再在 DNA 连接酶的催化下即可相互连接，形成人工重组体。常用体外重组方法主要有以下四种。

（一）黏性末端连接

目的基因和载体 DNA，经适当的同一个限制酶消化，使它们两端各产生相同的黏性末端。在适宜温度下，酶切后的目的基因片段和线性载体 DNA 按照碱基配对关系退火、互补碱基以氢键相结合，在 T_4DNA 连接酶的催化下，其末端以磷酸二酯键相连接，形成环状 DNA 重组体（见图 3－8）。该法因用同一个限制酶消化，产生完全相同的黏性末端，故称为全同源黏性末端连接法。该法最简单，而且连接效率高，在克隆繁殖后，用同一限制酶消化，可回收插入 DNA 片段。若将目的基因与载体 DNA 分别用两个不同的限制酶切割，使两种 DNA 各自片段 5′和 3′黏性末端不同，但两者配伍末端又互补，退火杂交、再经连接酶连接也可形成重组体，这种方法称非同源互补黏性末端连接法。

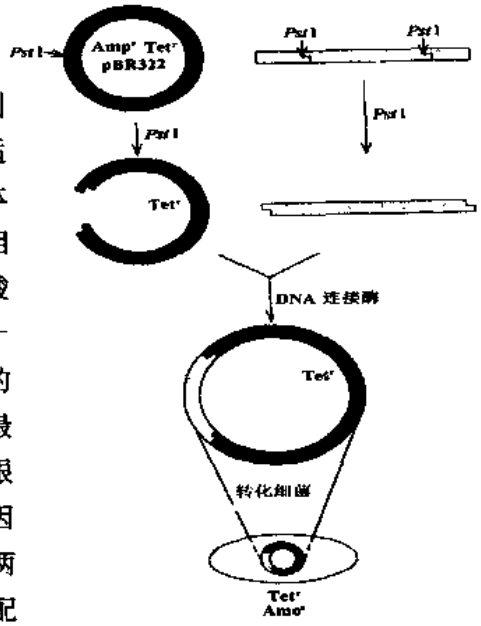


图 3－8 黏性末端连接

全同源黏性末端连接法，酶切后载体 DNA 两端的碱基序列互补，可能自身环化，形成空白载体，使 DNA 重组体的比例下降。避免这一缺陷的方法之一是，将载体

DNA 酶切后，用碱性磷酸酶除去 5′端磷酸基团，可防止载体 DNA 的自身环化。

（二）平端连接

利用 T_4DNA 连接酶可催化连接相同或不同限制酶切割的平端 dsDNA。平端连接的特点是可以恢复原有酶切位点，或产生一个新的酶切位点，这为重组体筛选鉴定提供了一条简捷的途径。其缺点是连接效率较低，为提高连接效率，在连接平端 dsDNA 片段时，所用的 ATP 及 T_4DNA 连接酶的浓度比黏性末端要高些。

（三）人工头连接

为提高平端连接的效率，可以使用人工接头（linker）法。所谓接头是指含有某些限制酶切位点的寡核苷酸片段。在 T_4DNA 连接酶的催化下，将接头接到目的 DNA 片段两端，再用相应的限制酶切割，这样目的 DNA 片段两端具有了黏性末端，可以连接到同一限制酶线性化了的载体上去。

（四）同聚物加尾连接

同聚物加尾连接是利用同聚物序列，如多聚 A 与多聚 T 之间的退火作用完成连接，该法是 cDNA 克隆常用的方法。在末端转移酶（terminal transferase）作用下，在 DNA 片段末端加上同聚物序列、制造出黏性末端，而后进行黏性末端连接。

三、重组 DNA 导入宿主细胞

外源 DNA 与载体 DNA 形成重组 DNA 后，需要导入宿主细胞才能复制、扩增与表达。在基因克隆技术中，将质粒 DNA 及其重组体导入细菌称为转化（transformation）；将病毒及其重组体导入宿主细胞称为转染（transfection）；将噬菌体及其重组体导入宿主细胞称为转导（transduction）。下面简单介绍几种常用方法。

（一）氯化钙法

将受体菌悬浮在低温 0℃和低渗 $CaCl_2$ 溶液中，经一段时间后，菌体细胞壁通透性增加，菌体膨胀成球形，易于吸收外源 DNA，这种状态的菌称谓感受态菌。将溶于磷酸盐缓冲液（PBS）中的重组质粒 DNA 加入感受态菌，搅拌混匀，此时转化液中的 DNA 形成抗 DNase 的羟基—磷酸复合物，黏附于细胞表面，经短暂的热体克（42℃90 秒）后即可进入细胞内。然后在不含抗生素的培养基中继续培养，使质粒 DNA 得到复制，并使抗生素抗性基因表达。随后将转化的细菌接种在含有有关抗生素的琼脂平板上，过夜生长，即可得到转化的菌落。此种方法转化频率很低（大约 0.01%～0.1%），故需一个很强的选择性标志基因。

（二）电穿孔法

电穿孔法最早用于 DNA 导入真核细胞，现已用于大肠杆菌和其他细菌。将对数生长期的大肠杆菌与外源 DNA 混合于电穿孔杯中，在高频电流作用下，细胞壁出现许多

小孔，外源 DNA 即可进入细胞内。电转化应在 0℃～4℃进行。该法的优点是无需制备感受态细胞，故操作较氯化钙法简单，其转化率也比较高，可获得 10^8～10^9 转化子/μgDNA。其缺点是需特殊仪器，电穿孔最适条件不易掌握，细胞杀伤率高（30%～70%）。

（三）脂质体转染法

近年相继发展了系列广谱和高效的脂质体，如单价阳离子脂质体 DOSTAP 等和多价阳离子脂质体 DOSPER 和 infect AMINE™等。其中以 infect AMINE™最新。脂质体是一类双层的微囊结构，如 DOSPER 和 infectAMINE™，含有一个亲水性正电荷的精胺基团头和疏水性脂酸尾，可与 DNA 分子结合成脂质体/多核苷酸复合物，使 DNA 免受 DNase 的降解，精胺基团可与带负电荷的细胞表面非特异吸附，易于通过细胞内吞作用而进入细胞内。该方法的优点是：转化率最高，转化细胞类型广泛。其缺点是脂质体价格昂贵。

（四）显微注射法

显微注射法是在显微镜下，经细胞玻璃针将外源 DNA 直接注入细胞。该方法是将 DNA 准确注入细胞核中最可靠的方法，特别是借助于电脑技术更为精确快速。该法适用于多种贴壁生长细胞及悬浮生长细胞，也是转基因植物和动物研究的最佳手段。其缺点是仪器昂贵，对实验者的操作技巧和熟练程度的要求比较高。

四、重组体的筛选与鉴定

在转化过程中并非每个宿主细胞都被转化；即使获得转化的细胞也并非都含有目的基因，可能含有自身成环的载体分子，或载体与非目的基因形成的重组子。因此，转化后须在不同层次、不同水平上进行筛选，以区别转化子与非转化子、重组子与非重组子，以及鉴定所需的特异性重组子。

（一）遗传学方法

1. 筛选转化菌　克隆载体上含有抗菌素抗药性基因，如 amp^r 和 tet^r 等用其转化细菌后，含有重组体的细菌能在含 amp 和/或 tet 的培养板上生长并形成菌落，不含有重组体的细菌则不能生长。利用这种方法即可把转化菌和非转化菌区分开，初步筛选出转化菌。

2. 筛选带有重组体的克隆

（1）插入失活法（insection inactivation）：许多载体带有抗菌素抗性基因，当目的基因插入某一抗药基因内，该基因就会失活，转化菌则会丧失此种抗药性。如 pBR322 带有 amp^r 和 tet^r 标记基因，tet^r 基因上含有 BamH Ⅰ位点，如果在该位点插入目的基因，则会造成 tet^r 基因失活。含该重组体的细胞能在含 Amp 的培养板上生长，而不能在 Tet 培养板上生长；若在 Amp 和 Tet 培养板上都能生长，则可能是载体自身环化后的转化子；未转化细胞在两种培养板上都不能生长。

（2）β—半乳糖苷酶显色反应法：　在 PUC 系列和 M13 中，均引入 *LacZ* 基因 α 肽编码序列，其编码产物为 β—半乳糖苷酶的 α 肽。当无外源 DNA 插入时，质粒表达 α 肽。突变型 Lac^- E. Coli 可表达该酶的 ω 肽。单独存在的 α 及 ω 肽均无 β—半乳糖苷酶活性，只有宿主细胞与克隆载体同时共表达这两个肽时，宿主细胞内才有 β—半乳糖苷酶活性，在含有底物 X—gal 和诱导剂 IPTG 条件下，菌落呈蓝色，这就是 α—互补（alpha complementation）。如果 *LacZ* 基因由于插入外源基因而失活，结果无 α 肽表达，转化菌落无 α—互补，缺乏 β—半乳糖苷酶活性，在 X—gal 培养板上为白色菌落，由于这种颜色标志，重组克隆和非重组克隆的区别一目了然，这种筛选也称为“蓝白筛选”，或称 α—互补筛选。

（二）限制性内切酶图谱鉴定

分别挑出单个菌落接种于培养基中，37℃培养过夜后，迅速提取质粒 DNA，经适当限制酶切割后，进行琼脂糖电泳，在紫外光下观察有无插入片段、插入片段大小和方向等。限制酶的选用必须根据载体和插入片段上的酶切位点来确定。

（三）核酸杂交法

将待选菌在琼脂平板上培养成菌落或噬菌斑，再把硝酸纤维素膜或尼龙膜盖于长有菌落或噬菌斑的琼脂平板的表面，定好位，把膜揭起，进行原位分子杂交（见图 3—9），根据阳性斑点在平板上的相应位置，即可找出阳性克隆。

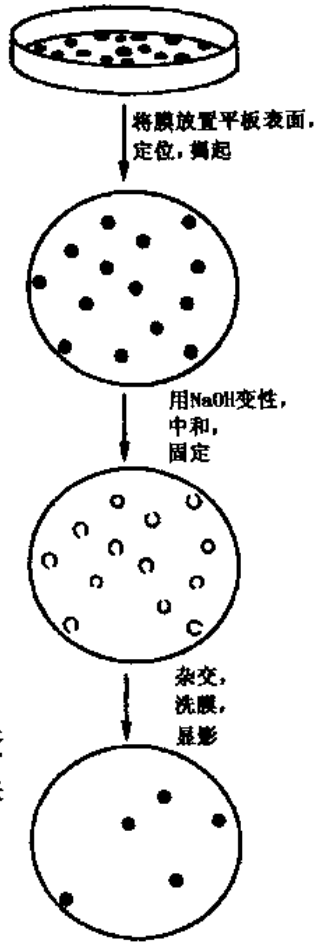


图 3—9　菌落原位杂交

（四）PCR 筛选法

一些克隆载体的外源 DNA 插入位点两侧存在已知序列，如 pGEM 载体系列的多克隆位点（MCS）两侧为 SP_6 及 T_7 启动子序列，设计并合成与 SP_6 及 T_7 启动子互补的 PCR 引物，对抽提的重组 DNA 进行 PCR 分析，不但可迅速扩增插入片段，而且可以直接进行 DNA 序列测定。

（五）免疫化学筛选法

如果外源基因能够在宿主细胞进行表达，合成外源蛋白质，可以采用免疫化学等方法检测，即用已知的抗体来检测目的基因编码的抗原。

第四节 真核基因在大肠杆菌中的表达

基因表达也是基因克隆的主要任务之一。根据被表达的基因和表达基因所用的系统可将表达分为四种：①原核基因在原核细胞中的表达；②真核基因在原核细胞中表达；③原核基因在真核细胞中表达；④真核基因在真核细胞中表达。目前，对真核基因在原核细胞中表达的研究较广泛、深入。本章以真核基因在大肠杆菌中的表达为例介绍克隆基因的表达。

一、真核基因在大肠杆菌中表达的基本条件

一种克隆的真核基因，在大肠杆菌中实现正确有效的表达需要考虑以下一些基本条件：①大肠杆菌没有剪切内含子的功能，所以，真核细胞的目的基因必须来自于 cDNA；②真核基因 mRNA 缺乏结合细菌核糖体的 SD 序列，因此 cDNA 的起始密码子（ATG）上游部分（5′端非编码区）是无用的，必须除去。对于一些分泌性蛋白，还应除去信号肽部分；③表达载体应是含有大肠杆菌 RNA 聚合酶所能识别的启动子（如 PL，tac，T_7等）和 SD 序列的大肠杆菌表达载体；④载体上只能用大肠杆菌启动子，将外源基因克隆在启动子下游，以便大肠杆菌 RNA 聚合酶识别启动子，并带动真核基因在大肠杆菌细胞中转录。目前，商用表达载体都含有启动子和 SD 序列，无需自己构建，只要选择合适的载体即可；⑤为提高表达效率，应根据启动子类型和特定的蛋白质，选择合适的菌株和诱导条件。

二、真核基因在大肠杆菌中的表达方式

真核基因在 E. coli 中可以表达为融合蛋白、非融合蛋白，或分泌型表达。

（一）融合蛋白

融合蛋白的氨基端是原核序列，羧基端是真核序列。融合蛋白中由于含有一段原核多肽序列，可能影响其真核蛋白的免疫原性，但也有其优点：①在菌体内比较稳定，容易实现高效表达；②基因操作比较简单；③在某些情况下，表达后的融合蛋白经化学处理或特异蛋白酶水解，可以切除其氨基端的原核多肽，而获得具有生物学活性的真核天然蛋白分子，同时避免了在表达产物氨基端带有起始密码子编码的甲硫氨酸。表达融合蛋白的必要条件是，将真核基因连接在一段原核基因的下游后，真核基因的读码框架必须与原核基因的读码框架相符合。

（二）非融合蛋白

非融合蛋白是指在细菌内表达的蛋白以真核蛋白 mRNA 的 AUG 为起始，在其氨基端不含任何原核多肽序列。为此，表达非融合蛋白的操纵子须改建成原核启动子—原核 SD 序列—真核基因的起始密码子、结构基因及终止密码子。其中 SD 序列与 AUG 之间的距离要适当，一般 3～12 个核苷酸为宜。可以通过改建一种表达载体，使其含有

原核启动子和 SD 序列，然后将真核基因的编码序列重组于 SD 序列下游 3～12bp 处，表达后即可获得真核蛋白。

（三）蛋白分泌型表达

将真核基因重组于编码原核蛋白信号肽序列的下游，以表达带有原核蛋白信号肽的融合蛋白。这种融合蛋白当分泌到位于 E. coli 细胞内膜与外模之间的周质时，其信号肽可被信号肽酶切割，而真核蛋白分泌至胞外。该表达方式有以下优点：①可减少所需蛋白质在宿主菌内的降解；②有些在细胞内表达时无活性的蛋白质分泌表达时则具有活性，原因是这些蛋白质能按适当、正确的方式折叠；③分泌后的蛋白质氨基端不含起始密码子编码的甲硫氨酸；④蛋白质分泌至胞外后便于提纯。因此，有些学者专门设计并构建了带有原核蛋白信号肽序列的分泌型表达载体。

（王守训）

第四章　定点诱变与蛋白质工程

对于存在于自然界的任何蛋白质来说，采用重组 DNA 技术克隆其基因，在特定宿主细胞中进行表达，可以获得运用于商业用途的纯化的产品。然而，这些蛋白质的理化特性常常不能适应工业用途。从生长于非常环境中的生物体中获得基因，可表达获得适用于特定目的的蛋白质。例如，从生存于 90℃温泉中的 *Bacillus stearothermophilus* 分离获得 α－淀粉酶基因，经表达可获得耐高温的 α－淀粉酶，该酶可用于从淀粉制造酒精的工业生产。传统诱变技术，也可以获得能编码具有期望特性的蛋白质的突变基因，并获得突变蛋白质。然而，在传统诱变方法中，经化学诱变剂或物理诱变剂处理的生物体，它的任何基因都有可能发生突变，而目的基因的突变频率又可能相当低，给突变体的筛选工作造成了很大的麻烦，而且所获得的突变蛋白质，常常因其氨基酸的改变而导致蛋白质活性下降，因而传统诱变方法实用价值不大。

随着分子生物学方法的进步，人们建立起通过改变克隆基因中的特定碱基，从而改造蛋白质的技术。即在体外，通过碱基取代、插入或缺失的方法，使基因 DNA 序列中的某个特定碱基发生改变（称为蛋白质工程）。通过蛋白质工程，人们可以随心所欲地改造蛋白质的结构，从而改变蛋白质的理化性质和生物学功能。例如，我们可以利用蛋白质工程改变酶的 K_m（Michaelis 常数）、V_{max}（酶催化反应的最大速度）值，酶促反应的最适温度、最适 pH 值、酶促反应的特异性以及酶蛋白的稳定性等。蛋白质工程与基因定点突变有着不解之缘，故本章在重点介绍定点诱变原理和技术的基础上，举例介绍一下蛋白质工程的应用。

第一节　定点诱变

随着分子生物学技术的发展，特别是基因克隆技术的应用，分离并研究单基因的结构与功能已成为一种常规的工作。与此相适应，基因诱变技术也有了极大的发展。现在，人们不仅能够对多细胞或是有机体作诱变处理，并从成千上万突变群体中筛选出期望的突变体，而且还能在体外试管中通过碱基取代、插入或缺失使基因 DNA 序列中任何一个特定的碱基发生改变。这种体外特异性改变某个碱基的技术，叫做定点诱变（site－directed mutagenesis）。由于它具有简单易行重复性高等优点，现已发展成为基因操作的一种技术。这种技术不仅适用于基因结构与功能的研究，还可通过改变基因的

密码子来改造天然蛋白质。目前已发展的定点诱变方法主要有寡核苷酸引物诱变、PCR诱变、盒式诱变及随机诱变等，下面将逐一予以讨论。

一、M13DNA 寡核苷酸诱变

(一) 原　理

基因工程中，噬菌体是一种常用的目的基因载体，噬菌体中又以 λ 和 M13 为最常用，它们都属于大肠杆菌噬菌体（coliphage）。M13 噬菌体是一种环型 DNA，分子量为 6.4Kb，如同其他单链 DNA（single stranded DNA，ssDNA）噬菌体情况一样，在颗粒中包装的仅是正链的 DNA，有时也称为感染型单链（图 3－6）。当感染型单链 M13 噬菌体感染大肠杆菌后，在菌体内借助宿主的酶系统先把 ssDNA 复制为双链（dsDNA），称为复制型（replication form，RF）M13（RF－ M13）。RF－ M13 在菌体内继续繁殖至每个菌体达 100－200 个拷贝后，转变为合成不对称，只产生单链 DNA，然后再从感染的宿主菌中分泌出来，并再度感染其他宿主细胞。胞内的 RF 型分正负链，在胞外正负链均可作模板进行半保留复制。感染型的单链以正链为主。

M13 感染大肠杆菌后，进行混合培养。培养液经离心沉淀，从其上清液收获到 M13 的单链 DNA 可作为复制模板。将培养液经离心沉淀得到的细胞沉渣用化学方法破碎细菌，按照质粒提取的方法可制备 RF 型双链 DNA。M13RF－DNA 可接纳目的基因，组成重组体。含目的基因的 M13 重组体，又可再感染或转化大肠杆菌，提取带目的基因的单链 DNA，可作目的基因的定点诱变。

寡核苷酸定点诱变技术所依据的原理是按照体外 DNA 重组技术，将待诱变的目的基因插入到 M13 噬菌体上，制备此种含有目的基因的 M13 单链 DNA，即正链 DNA。再使用化学合成的含有突变碱基的寡核苷酸短片段作引物，启动单链 DNA 分子进行复制，随后这段寡核苷酸引物便成为新合成的 DNA 子链的一个组成部分。因此所产生的新链便具有已发生突变的碱基序列。为了使目的基因的特定位点发生突变，所设计的寡核苷酸引物的序列除了所需的突变碱基外，其余的则与目的基因编码链的特定区段完全互补。

(二) 诱变过程

M13 寡核苷酸诱变过程的主要步骤如下（见图 4－1）：

1. 正链 DNA 的合成　将目的基因克隆到 M13 噬菌体中，制备含有目的基因的 M13 单链 DNA，即正链 DNA。

2. 突变引物的合成　应用化学法合成带错配碱基的诱变剂寡核苷酸片段，即寡核苷酸引物，其中除了含有特殊的突变碱基外，其他碱基与目的 DNA 的适当区域互补。

3. 异源双链 DNA 分子的制备　将突变引物 DNA 与含目的基因的 M13 单链 DNA 混合退火，使引物与待诱变核苷酸部位及其附近形成一小段具有碱基错配的异源双链的 DNA。在 Klenow 片段催化下，引物链便以 M13 单链 DNA 为模板继续延长，直至合成全长的互补链，然后再由 T_4DNA 连接酶封闭缺口，最终在体外合成出闭环异源双链的

M13DNA 分子。

4. 闭环异源双链 DNA 分子的富集和转化

因为在体外合成异源双链的 M13DNA 分子后，尚余有单链 M13 噬菌体 DNA 或具裂口的双链 M13DNA 分子，转化大肠杆菌后，也会增殖而产生很高的转化本底。故转化前应使用 S_1 核酸酶处理法，或碱性蔗糖梯度离心法，减少本底，使闭环的异源双链的 M13DNA 分子得到富集。然后将富集的闭环异源双链 M13DNA 分子转化给大肠杆菌细胞后，产生出同源双链 DNA 分子。

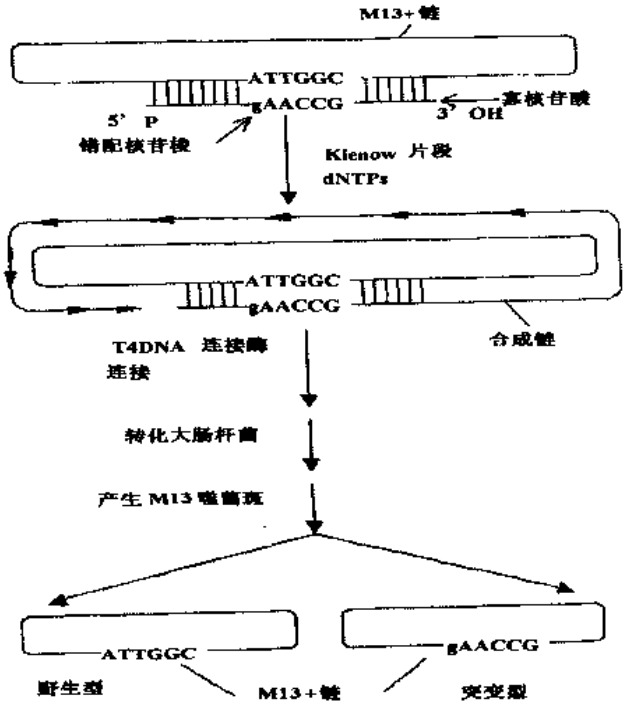


图 4－1 M13 寡核苷酸诱变过程示意图

5. 突变体的筛选 闭环异源双链 DNA 分子转化大肠杆菌后可产生野生型和突变型两种转化子，二者混合存在，故需进行筛选以获得突变型转化子。常用筛选方法有链终止序列分析法、限制位点法、杂交筛选法和生物学筛选法。其中杂交筛选法最简单也最有用。在杂交实验中，以诱变剂寡核苷酸为探针，在不同温度下进行噬菌体斑杂交，选择突变体克隆。由于探针与野生型 DNA 之间存在着碱基错配，而与突变型则完全互补，于是便可以根据两者杂交稳定性的差异，筛选出突变型的噬菌斑。

6. 突变基因的鉴定 对突变体 DNA 作序列分析，检测突变体的序列结构特点，有助于确定在诱变过程中是否引入其他偶然错配。

（二）Kunkel 定点诱变法

体外 DNA 合成往往是不完全的，所以部分合成的 DNA 分子必须通过蔗糖密度梯度离心除去。理论上来说，DNA 是半保留复制的，应用寡核苷酸定点诱变时，所形成的噬菌体中携带突变基因的应为一半。但实际上，由于技术上的原因通常只有 1%～5%的噬菌斑含有突变基因的噬菌体。因此，为了获得更多含有突变噬菌体的噬菌斑，必须提高突变体的比率。目前已有多种改良的寡核苷酸定点诱变方法，此处将 Kunkel 1985 年建立的方法做以简单介绍。

Kunkel 定点诱变法是一种通过筛除含有尿嘧啶的 DNA 模板链进行的寡核苷酸定点诱变法。它的基本原理是，将待突变的基因克隆入 RF－M13 DNA 载体上，导入具有 dUTP 酶（dut）和 N－尿嘧啶脱糖苷酶（ung）双缺陷的大肠杆菌（dut^-，ung^-）菌株中。Dut 缺陷导致细胞内 dUTP 水平上升，并在 DNA 复制时，部分取代 dTTP 进入 DNA 新生链中。又由于 ung 缺陷使掺入 DNA 的 dUTP 残基不能除去。由这种大肠

杆菌菌株产生的 M13 单链 DNA 大约有 1%的 T 被 U 所取代，然后进行 DNA 另一条链的体外合成，双链 DNA 导入正常的大肠杆菌中，其尿嘧啶 N－糖基化酶除去 DNA 链上的尿嘧啶碱基。结果原来的 M13 模板链被降解。只有突变链因不含 U，被保留下来（见图 4－2）。这种方法产生的 M13 噬菌体中含有突变 DNA 的比例大大增加。

二、PCR 定点诱变

寡核苷酸诱变方法所使用的 M13 噬菌体系统，操作起来比较繁琐，诱变周期较长。随着 PCR 技术的不断发展，人们将其应用在诱变技术中，使得此技术得以简化。

1. PCR 的基本原理　聚合酶链反应（polymerase chain reaction，PCR）或称体外 DNA 扩增技术，1985 年由美国 Cetus 公司的 Kary Mullis 首创，是一种极为简便和快速的体外 DNA 扩增技术，能在短时间内将数量上仅为几个拷贝的基因 DNA 放大数百万倍，大大简化了传统的分子克隆技术。PCR 原理类似于天然 DNA 的复制，主要是利用 DNA 聚合酶依赖于 DNA 模板的特性模仿体内的复制过程。PCR 技术的特异性决定

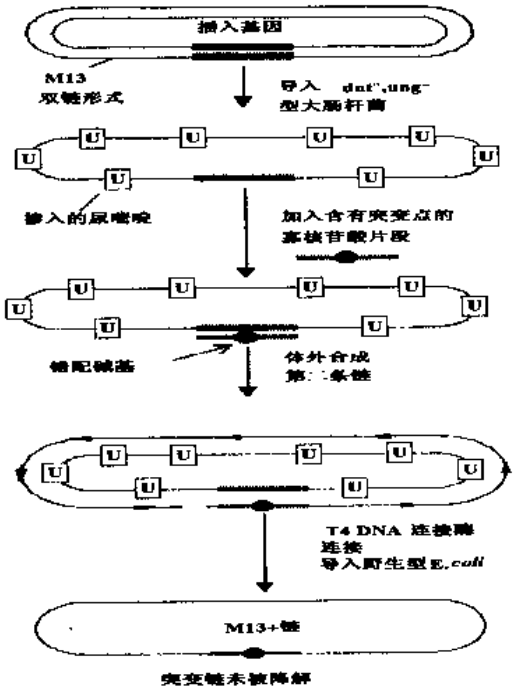


图 4－2　Kunkel 定点诱变示意图

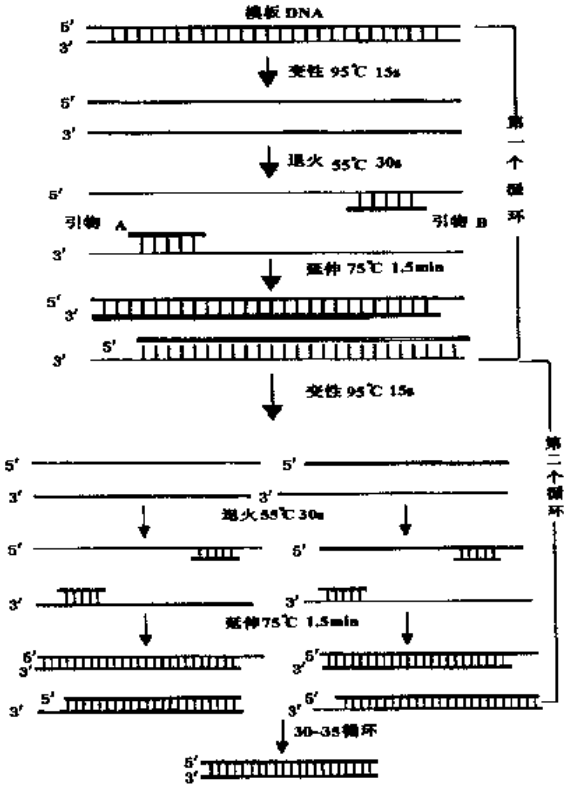


图 4－3　PCR 原理示意图

于引物和模板 DNA 结合的专一性。其基本过程（见图 4－3）如下：①变性（denaturation）：通过加热使模板 DNA 双螺旋的氢键断裂，双链解开形成单链。一般为 95℃，15sec；②退火（annealling）：当温度突然降低时，人工合成的两个寡核苷酸引物分别与模板单链形成杂交链。一般为 55℃、30sec；③延伸（extension）：将反应系统调节到酶的最适温度，以四种 dNTP 为底物，DNA 聚合酶催化以引物为起始点的 DNA 链的延伸反应，一般为 72℃，1.5min。以上三步为一个循环，每一循环的产物可

作为下一循环的模板，因此经数小时 n 次循环后，目的 DNA 理论上可扩增 2^n 倍。常规 PCR 反应采用 30～35 个循环，最后一次循环的延伸时间一般加长至 5min，以获得更多的扩增产物。反应体系中引物的量远远大于模板 DNA；另外模板 DNA 链长，致使引物与模板杂交率大于模板 DNA 的自主互补。

2. PCR 诱变　PCR 定点诱变又称为 PCR 寡核苷酸定点诱变，该法具有简单、快捷的特点，其基本原理及操作程序如下（图 4－4）：①将待诱变靶基因克隆到质粒载体上，并分装到两个反应管中；②在每一个反应管中加入两种特定的引物，其中引物 1 和 3 均含有错配核苷酸，但两个引物分别与质粒 DNA 的不同链不完全互补，引物 2 和 4 均不含有错配核苷酸，二者分别与质粒 DNA 的引物 1 和 2 杂交链的互补链完全互补；③进行 PCR 扩增获得含有突变碱基的线型质粒 DNA；④将两个反应管中的线型质粒 DNA 混合，再经过变性和复性，一个反应管中的一条链和另一个反映管中的互补链杂交，通过两个粘性末端形成带有缺口的环状 DNA 分子；⑤转化入大肠杆菌，环状 DNA 分子的缺口可被大肠杆菌修复。如果同一反应管中的两个互补链又互相杂交，则继续形成线状 DNA 分子，在大肠杆菌中不稳定，易被降解。该方法把特异突变点导入克隆基因，无需把基因插入 M13 中，即可在大肠杆菌中进行表达。

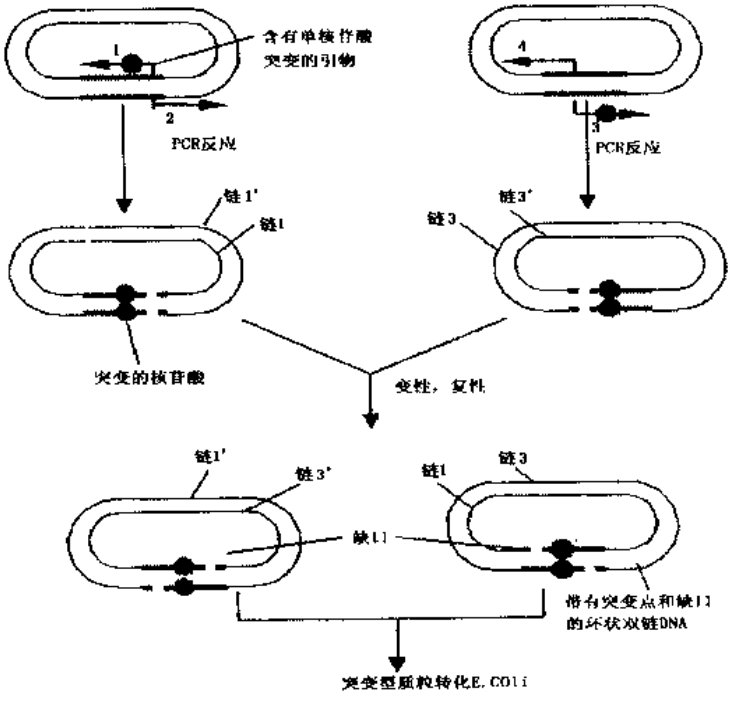


图 4－4　PCR 进行寡核苷酸定点诱变的示意图

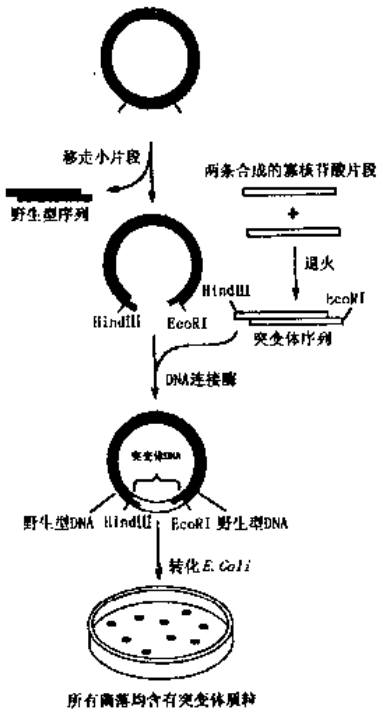


图 4－5　盒式诱变示意图

三、盒式诱变

盒式诱变（cassette mutagenesis）是一种定点诱变技术，其方法是利用一段人工合成的具有突变序列的双链寡核苷酸片段，取代野生型基因中相应序列（见图 4－5）。这种诱变的双链寡核苷酸片段是由两条人工合成的寡核苷酸链组成的，当它们退火时，会

按照设计要求产生出克隆需要的粘性末端。这些合成的寡核苷酸片段就好像不同的盒式录音磁带，可随时插入到已制备好的载体分子（“录音机”）上，便可以获得数量众多的突变体，故称为盒式诱变。该方法的优点是简单易行，突变效率高。

四、随机诱变

定位诱变通常需要了解目的序列的详细情况，当缺乏这方面的资料和信息时，定位诱变方法的利用就受到限制。在这种情况下，利用随机诱变方法，在目的序列中产生突变，仍可以用于研究目的蛋白或目的核酸序列的结构和功能。

随机诱变（random mutagenesis）的工作原理（见图 4－6）是，将待突变基因克隆在一个载体的特定位点上，其下游紧接着是两个限制性核酸内切酶的酶切位点（RE1，RE2），前一酶切位点是 5′突出（3′凹陷）的单链末端，后一酶切位点是 3′突出（5′凹陷）的单链末端。然后用大肠杆菌核酸外切酶Ⅲ（exonucleaseIII，ExoIII）处理酶切缺口。EcoIII 的主要活性是催化双链 DNA 自 3′－羟基端逐一释放 5′单核苷酸，其底物是线性双链 DNA 或有缺口的环状 DNA，而不能降解单链或双链 DNA 的 3′突出末端。因此，当用 ExoIII 处理时，则可逐一水解 3′凹陷末端。在适当时机，终止 ExoIII 的酶切反应，缺口用 Klenow 补平，底物为 4 种 dNTP，再加上一种脱氧核苷酸的类似物。在缺口填补过程中，这个类似物会掺入到 DNA 链上的一处或多处。再用 S1 核酸酶处理单链末端，形成平头末端，并用 T_4DNA 连接酶连接。这种重组质粒转化大肠杆菌后，50％的基因上携带有错配的碱基，导致位点变异。

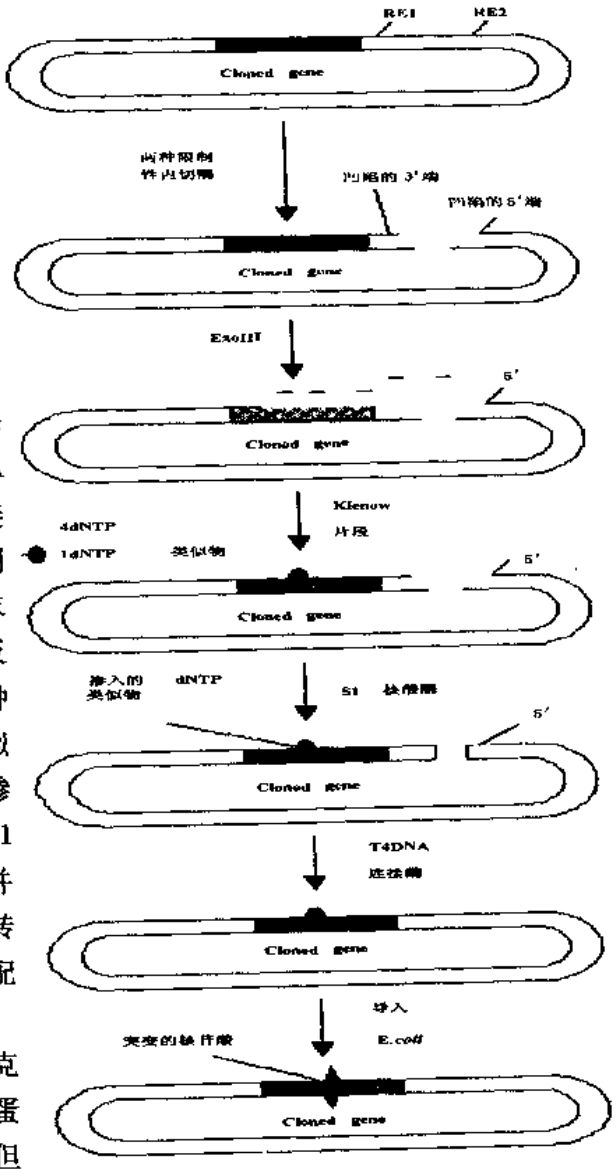


图 4－6　随机诱变示意图

随机诱变的缺点是必须检测每个克隆，看看哪一个产生了具有期望特性的蛋白质。这种检测不是一个简单的工作，但这是发现有新特性蛋白质的唯一方法。一旦发现了有潜在优点的突变体，通过确定哪个位点突变了可以确定克隆基因的序列。

第二节 蛋白质工程

随着基因定点诱变和基因克隆技术的发展，人们可以根据设计特异地改变一个基因编码的氨基酸序列，产生具有新的特性的蛋白质，以适应工业化生产的需要，使得蛋白质工程得以迅速发展。蛋白质工程的主要研究内容是，通过特异性的改造蛋白质中氨基酸的序列，提高蛋白质的生物学功能，提高其在酸、碱、热等极端条件下的稳定性，或者使其便于分离纯化等。此处对蛋白质工程的设计思想作以简单介绍。

一、加入二硫键

蛋白质分子中两个半胱氨酸残基上的巯基（—SH）之间氧化脱氢即形成二硫键（—S—S—）。二硫键的数量与蛋白质的稳定性有关，增加蛋白质分子中的二硫键，可提高蛋白质分子的热稳定性、有机溶剂稳定性和酸碱稳定性。在一项研究中，通过寡核苷酸定点诱变构建了 T4 溶菌酶的六种变异体（见表 4—1），其中 3 个变异体分别将 2 个氨基酸残基诱变为半胱氨酸，使分子内增加了 1 个二硫键，有 2 个变异体增加了 2 个二硫键，1 个变异体增加了 3 个二硫键。新介入的半胱氨酸在酶蛋白中的位置 3 和 97，9 和 164，21 和 142 之间形成二硫键。诱变后的基因，在大肠杆菌中表达，纯化后分别测定酶活性和热稳定性。蛋白质的热稳定性常常被定义为蛋白质的全部结构有 50%变性的温度。T4 溶菌酶的野生型（自然的）酶构型中有两个半胱氨酸，但是都没有形成二硫键。在所谓的假野生型酶中，通过寡核苷酸定点诱变，这些半胱氨酸在没有改变酶的活性和热稳定性的情况下，被改变成苏氨酸和丙氨酸。因此，这个假野生型序列提供了一个比较具有潜在热稳定性的二硫键的变体的标准，而且预防在突变的半胱氨酸残基和自然发生的半胱氨酸残基之间形成假的二硫键。

表 4—1 T4 溶菌酶和 6 种设计的变体特性

酶	氨基酸的位置							二硫键的数目	相对活性	Tm
	3	9	21	54	97	142	164		(%)	(℃)
wtα	Ile	Ile	Thr	Cys	Cys	Thr	Leu	0	100	41.9
pwt	Ile	Ile	Thr	Thr	Ala	Thr	Leu	0	100	41.9
A	Cys	Ile	Thr	Thr	Cys	Thr	Leu	1	96	46.7
B	Ile	Cys	Thr	Thr	Ala	Thr	Cys	1	106	48.3
C	Ile	Ile	Cys	Thr	Ala	Cys	Leu	1	0	52.9
D	Cys	Cys	Thr	Thr	Cys	Thr	Cys	2	95	57.6
E	Ile	Cys	Cys	Thr	Ala	Cys	Cys	2	0	58.9
F	Cys	Cys	Cys	Thr	Cys	Cys	Cys	3	0	65.9

wtα：野生型 T4 溶菌酶；pwt：假野生型酶；A—F：六种设计的半胱氨酸变体；Tm：熔点温度

这个实验的结果表明酶的热稳定性升高是二硫键存在的结果，加入二硫键的数量与

蛋白质热稳定性的增加呈线性关系。但是某些比野生型或者假野生型有更好的热稳定性的变体（C，E 和 F）已经失去了它们的酶活性。三个变体酶活性的丧失可能反映了在 21 和 142 位残基之间含有一个二硫键连接的多肽分子骨架的扭曲变形。在许多情况下，蛋白质工程是一个“实验和失败”的过程。因此，能产生“最好的”变体的准确的氨基酸的改变总是不能被预知的。

表 4－2　几种工业用酶及其应用

酶	工业应用
α淀粉酶	啤酒和酒精的生产
氨酰酶	L－氨基酸的制备
菠萝蛋白酶	肉制品嫩化剂，果汁净化
过氧化氢酶	制备食品的抗氧化剂
纤维素酶	酒精和葡萄糖的生产
无花果［蛋白］酶	肉制品嫩化剂，果汁净化
葡萄糖淀粉酶	生产啤酒和酒精
葡萄糖异构酶	高级果糖糖浆的生产
葡萄糖氧化酶	制备食品的抗氧化剂
蔗糖酶	蔗糖转化
乳糖酶	乳清的应用，乳糖水解
脂肪酶	生产奶酪，制备调味品
木瓜蛋白酶	肉制品嫩化剂，果汁净化
果胶酶	果汁净化，酒精生产
蛋白酶	去污剂，酒精生产
粗制凝乳酶	制备奶酪

在有活性的酶中，打算改变成半胱氨酸残基的氨基酸在空间上要相互靠近。这种空间上的邻近能确保整个分子构型将会基本上不受新的二硫键连接形成的影响。目标氨基酸和酶的活性位点没有关系。所谓活性位点就是酶分子中可能对构型的小的变化最敏感的部分。增加功能蛋白的稳定性，延长其体内半衰期，并抗酸性环境，对基因工程药物的开发具有重要意义。如目前正在开发或已投入上市的急性心肌梗死特效溶栓药物－组织血纤维蛋白溶酶原激活因子（t－PA）和尿激酶原（pro－UK），均为基因工程产品，在体内半衰期为 15 分钟左右，注射给药。若能提高其对酸的稳定性及对蛋白酶的抗性，那么这两种药物不但可用于心肌梗死病人的急救，还可口服用于心血管疾病的预防，其市场可以扩大 10 倍。表 4－2 列举了几种工业用酶及其应用。

二、将 Asn 和 Gln 转换成其他氨基酸

当蛋白质暴露于高温时，天冬酰胺（Asn）和谷氨酰胺（Gln）残基可能发生脱氨，这个反应释放出氨。随着酰胺部分地丧失，这些氨基酸分别变成了天冬氨酸和谷氨酸。肽链折叠的局部的改变可能会随之导致活性的丧失。

酵母的丙糖磷酸异构酶被用于研究天门冬酰胺的改造实验，该酶是由两个相同的亚基组成的二聚体，每个亚基都含有两个 Asn 残基，可能与该酶的热稳定性有关，因为四个 Asn 均位于两个亚基相互接触的表面上。通过寡核苷酸定点诱变技术，将第 14 和第 78 位上的 Asn 转换成 Asp，则变体酶的热稳定性下降，而将 2 个 Asn 分别换为 Thr 和 Ile，其半衰期则延长（见表 4－3）；若将两个 Asn 全部换成 Asp，则变体酶即使在常温下也不稳定，而且酶活性也大为降低。

表 4－3 酵母菌磷酸丙糖异构酶及其变体的热稳定性

酶	氨基酸位点		半衰期（min）
	14	78	
野生型	Asn	Asn	13
变体 A	Asn	Thr	17
变体 B	Asn	Ile	16
变体 C	Thr	Ile	25
变体 D	Asp	Asn	11

酶的稳定性以在 100℃是半衰期或者失活率来表示。半衰期越长酶越稳定

三、减少游离 Cys 残基的数目

在利用 DNA 重组技术表达外源基因中，常常会遇到这种情况：外源蛋白的表达量很大，但其生物活性却大大降低，即外源蛋白的比活很低，这很可能是由于在宿主细胞内，分子伴娘（chaperon）的缺失而引起的外源蛋白的无规则折叠所致，利用蛋白质工程技术可以减少游离的半胱氨酸（Cys）残基数目，以减少蛋白错误折叠的可能性，从而提高蛋白质的生物活性，这方面的例子是：当人 β－干扰素 cDNA 基因（β－IFN）在大肠杆菌中表达时，表达出的蛋白质生物活性只及天然糖基化蛋白质的 10%，尽管在大肠杆菌中 β－IFN 表达量很高，但蛋白质绝大部分是二聚体，单聚体含量很低，β－IFN DNA 序列分析显示，所编码的蛋白质有三个 Cys 残基，这些 Cys 可能涉及 β－IFN 分子间二硫键的形成，导致二聚体甚至多聚体的产生，这种情况只出现在大肠杆菌中而不发生在人体中，于是便产生了一个想法，就是将 β－IFN 分子中的 Cys 残基一个或多个用 Ser 来取代，因为 Cys 与 Ser 分子结构相似，前者有一个 S 原子，而后者是一个 O 原子，那么，三个 Cys 残基究竟应转变哪一个呢？有关 β－IFN 的信息不全，因为 α－IFN 与 β－IFN 同源，所以人们便借助于 α－IFN 进行推测。

α－IFN 分子中共有 4 个 Cys 残基，分别形成两对二硫键，经过比较有理由认为 β

－IFN 分子中第 17 位 Cys 残基是形成分子间二硫键的来源。因此取代的目标应是第 17 位的这个 Cys。这个推断被实验证明是正确的。诱变了的 β－IFN（第 17 位 Ser）在大肠杆菌中表达后，不再形成二聚体和多聚体。而且，Ser－17β－IFN 的生物比活性与天然的 β－IFN 相似，并且比天然蛋白质更稳定。

四、增加酶的活性

利用定点诱变技术不仅可以提高蛋白质的稳定性，还可以提高酶的活性。要改变酶的活性，需要一个详细描述了酶的活性位点的图谱。有了这样的资料，研究者们即可推测酶与底物的亲和程度，并利用定点诱变技术对蛋白质进行改造，提高酶的活性。

嗜热芽孢杆菌的酪氨酰－tRNA 合成酶通过两步反应催化酪氨酸和 tRNA 的结合。在第一步反应中，酪氨腺嘌呤核苷酸被 ATP 激活产生酶结合酪氨腺嘌呤核苷酸，伴随焦磷酸盐形成。在第二步中，酪氨腺嘌呤核苷酸被随后的 tRNA 的游离的 3’－羟基水解，因此酪氨酸部分被连接在 tRNA 上，并释放出 AMP。这两个反应都发生于当底物和酪氨酰－tRNA 结合酶结合时。

来自嗜热芽孢杆菌的酪氨酰－tRNA 合成酶的三维结构已经被确定，并且其活性位点已经被用图谱标出。因此，借助于计算机成像，可以预测改变活性位点的一个或多个氨基酸残基，对于酶和底物之间的相互作用会产生什么样的影响。为了检测这些是否正确，采用寡核苷酸定点诱变对酪氨酰－tRNA 合成酶的基因进行了特异性的修饰。第 51 位的苏氨酸（Thr）被用一个丙氨酸（Ala）或者脯氨酸（Pro）代替。在天然的酶中，第 51 位的苏氨酸的羟基基团与酪氨酸腺嘌呤核苷酸的核糖部分的环氧基形成一个长的氢键。据推测，去掉这个弱的氢键可能会改善酶与 ATP 的亲和力。

通过测定它们的动力学常数来确定这些改变了的酶的变体的特性，某些观察到的变化比预测的更加激动人心（见表 4－4）。Ala－51 变体酶和 ATP 的亲和力升高了大约两倍，而反应速度则无显著改变；Pro－51 变体酶对 ATP 的亲和力则提高了近 100 倍，反应速度也大大提高。

表 4－4　天然和诱变酪氨酰-tRNA 合成酶活性

酶	K*cat* (s^{-1})	Km (mmol/L)	K*cat*/Km ($s^{-1}mol^{-1}$)
Thr－51	4.7	2.5	1.860
Ala－51	4.0	1.2	3.200
Pro－51	1.8	0.019	95.800

五、改变酶的特异性

通过蛋白质工程技术还可以改变酶催化反应的特异性。现以葡萄球菌核酸酶为例予以简单介绍。

葡萄球菌的核酸酶，酶的专一性相对不高。它是一个 149 个氨基酸的单一多肽链，

催化降解单链 RNA 和单双链 DNA 的磷酸二酯键，作用底物位点是 AU 或 AT 富集区，产生 3’一磷酸和 5’一羟基末端，根据 X－晶体衍射结果得到的详细结构数据表明，将一小段已知序列的寡核苷酸片段，接在靠近酶催化活性位点的某一氨基酸残基上，可以提高该酶的水解特异性。

利用寡核苷酸定点诱变技术将天然核酸酶的 Lys－116 变成 Cys－116，然后将人工合成的寡聚核苷酸加到修饰酶中，这个寡聚核苷酸的 3’端带有一个 SH 基，这样便能在寡聚核苷酸与修饰酶之间形成一对二硫键（见图 4－7）。天然葡萄球菌核酸酶分子中没有 Cys 基。预测，在寡聚核苷酸片段与修饰酶的混合物中可能出现三种情况的二硫键：①核酸酶与核酸酶之间的二硫键二聚体；②寡聚核苷酸与寡聚核苷酸之间的二聚体；③核酸酶与寡聚核苷酸之间的混合二硫键。反应混合物可用凝胶过滤方法根据分子量不同将三种物质分开。

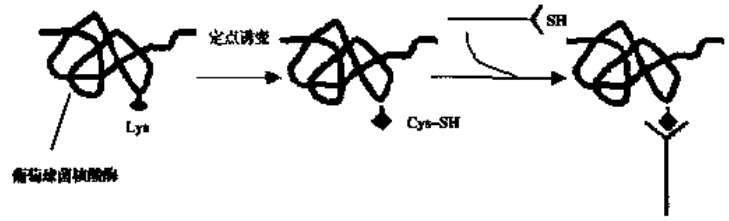


图 4－7 诱变葡萄球菌核酸酶与含-SH 寡核苷酸结合示意图

当携带有寡核苷酸片段的葡萄球菌修饰核酸酶用于单链 DNA 水解时，只有那些含有与工程酶分子上寡核苷酸片段互补区域的 DNA 分子，才能在其互补区域邻近的地方降解。于是这个工程酶变成了一个 DNA 序列专一的核酸酶。

（王伟　刘贤锡）

第五章　酶蛋白质工程

酶（enzyme）的应用已广泛渗透到工业、农业和医学领域。由于生物样品中酶的来源有限，又由于酶尚存在不稳定性、专一性不理想等缺点，人们一直期待着能按照人们的意愿改造酶，或制造酶的类似物，从而扩展酶的开发应用。基因工程和蛋白质工程技术的相继兴起，给改造酶，甚至创建新型酶提供了新的途径。本章拟在简单介绍酶学基本理论的基础上，介绍几种基因工程酶的有关内容。

第一节　酶的分子结构与催化功能

酶是一类生物催化剂。经典酶是指由活细胞合成的，对其特异底物起高效催化作用的一类特殊蛋白质。1982 年 Thomas Cech 首先发现 rRNA 前体具有自身催化作用，并提出了核酶的概念。本章内容主要是酶蛋白质工程，故以下仅介绍经典酶的结构与功能。

一、酶的结构特点与功能

（一）酶的化学本质

经典酶的概念认为酶的化学本质是蛋白质，这是 1926 年美国生化学家 James B. Sumner 第一次从刀豆提纯脲酶结晶，并证明其蛋白质本质后提出来的，得到生化界的公认。有的酶是单链多肽，如胃蛋白酶、胰蛋白酶等，称为单体酶。有的酶是多聚体，由数个相同或不同的亚基组成，称为寡聚酶。寡聚酶分子中有的亚基具有催化作用，称为催化亚基；有的亚基具有调节作用，称为调节亚基。不同的酶聚合在一起，完成一个化学反应，称为多酶体系，如丙酮酸脱氢酶、硫辛酸乙酰基转移酶、二氢硫辛酸脱氢酶等聚合成多酶体系，共同完成丙酮酸氧化脱羧生成乙酰辅酶 A 的反应，称为丙酮酸脱氢酶系。

（二）酶的分子组成

有的酶只含有蛋白质，也即水解产物中只含有氨基酸，这些酶称为单纯蛋白质酶类（simple enzymes），如蛋白水解酶类、淀粉酶和核糖核酸酶等。有的酶由蛋白质部分和

非蛋白质部分共同构成，称为结合蛋白质酶类（conjugated enzyme 结合酶）。其中蛋白质部分称为酶蛋白，非蛋白质部分称为辅酶或辅基，总称辅因子。酶蛋白决定酶的特异性，辅因子参与酶活性中心的组成，起传递化学基团或电子作用。酶蛋白和辅因子组合而成全酶，只有全酶才具有催化活性。

（三）酶的活性中心

1. 活性中心和必需基团的概念　酶分子中存在的各种化学基团并不一定都与酶的活性有关。其中那些与酶的活性密切相关的基团称为酶的必需基团（essential group）。这些必需基团在一级结构上并不互相毗邻，往往分散在氨基酸序列中，甚至分布在不同的肽链上。当肽链盘曲、折叠形成空间结构时，互相隔离的必需基团彼此靠近，集中在酶分子表面而形成具有三维结构的特定区域，该区域能与底物结合并发挥催化作用，故称为酶的活性中心（active center）或活性部位（active site）。对结合酶来说，辅酶或辅基参与酶活性中心的组成。酶活性中心的基团中，有的担负与底物结合的作用，使底物与酶形成复合物，称为结合基团；有的能影响底物中某些化学键的稳定性，催化底物转变成产物，称为催化基团。有时结合基团也具有催化作用。还有一些必需基团虽然不参加活性中心的组成，但却为维持酶活性中心应有的空间构象所必需，这些基团称为活性中心外的必需基团（见图 5－1）。酶活性中心中常见的必需基团有组氨酸残基的咪唑基、丝氨酸残基的羟基、半胱氨酸残基的巯基以及谷氨酸残基的 γ－羧基等。酶活性中心的三维结构区域或为裂缝，或为凹陷。此裂缝或凹陷多为氨基酸残基的疏水基团组成的疏水环境，形成疏水“口袋”。当底物进入疏水口袋后，多借助于次级键，如氢键、离子键、疏水键、Vander Waals 力等与酶结合并被催化转变为产物。不同酶的活性中心的构象不同，这是酶催化作用具有高度特异性的分子基础。

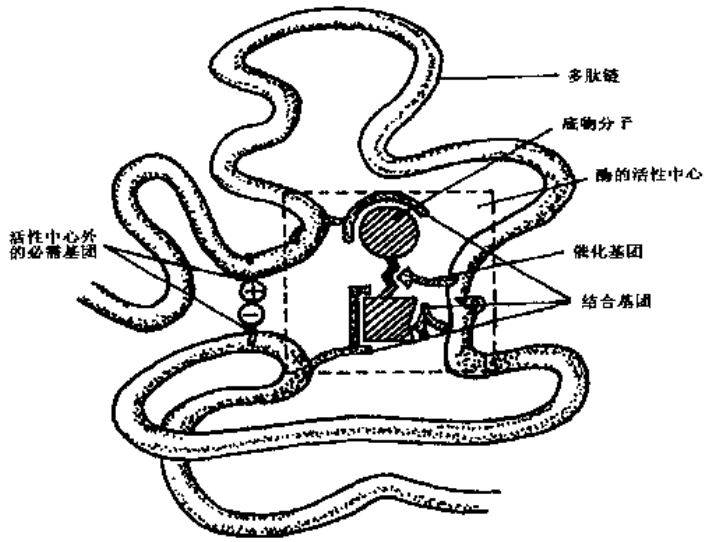


图 5－1　酶的活性中心示意图

有些酶在初合成时尚不具有酶的活性称为酶原。酶原之所以不具有催化活性是因为其必需基团分散存在于多肽链的氨基酸顺序中，相距较远，未能形成活性中心。在激活剂的作用下，必需基团相互靠近，构成活性中心，无活性的酶原转变成为有活性的酶，这种现象称为酶原激活。如胰凝乳蛋白酶活性中心中有 Ile16，His57，Asp102，Asp194 和 Ser195（序号以酶原为准），它们在酶原形成时分散在一条肽链上，但酶原经胰蛋白酶激活后，形成新的空间结构，必需基团相互接近，形成活性中心，具有了催化活性。由此可见酶的活性中心不是特异的氨基酸顺序，而是特异的氨

基酸组。

2. 活性中心与酶的专一性　酶催化反应的中间产物学说认为，酶在发挥催化作用之前，必须先与底物结合，形成中间复合物。酶与底物的结合要求底物具有特定的结构，即酶对其所催化的底物具有较严格的选择性，这就是酶的特异性或称专一性。(specificity)。根据酶对其底物结构选择的严格程度不同，酶的专一性大致可分为如下几型：①相对专一性，这类酶对底物结构的选择性相对较差。如酯酶只要求其水解的键是酯键，对酯键两侧的基团并无要求，此类专一性称为键专一性。有的酶除了需要正确的化学键，还要求键一侧的基团有特定的结构，如胰蛋白酶只能水解由碱性氨基酸（Lys 或 Arg）的羧基形成的肽键，这种专一性称为基团专一性；②绝对专一性，此类酶只能作用于特定结构的底物，进行一种专一的反应，生成一种特定结构的产物，例如脲酶仅能催化尿素水解成 CO_2 和 NH_3，而不作用于一切尿素的衍生物；③立体异构专一性，此类酶仅作用于立体异构体中的一种，如乳酸脱氢酶仅能催化 L－乳酸脱氢，而不作用于 D－乳酸；④序列专一性，这类专一性主要存在于大分子底物中，包括核酸、蛋白质和聚糖等。如限制性核酸内切酶能水解磷酸二酯键，这类酶不但对组成该磷酸二酯键两侧的脱氧核苷酸有严格的专一性，还要求该键临近的脱氧核苷酸有一定的序列，故暂命名为序列专一性。这类酶一般可识别 4－6 个连续的核苷酸序列，改变一个就不能作用，这是利用限制性核酸内切酶来研究基因 DNA 点突变的酶学基础。

3. 酶活性中心化学基团的鉴定　了解参与酶活性中心的各种化学基团的性质，研究酶结构与功能的关系，对酶工程学研究均具有十分重要的意义。

（1）物理学方法：用 X 射线衍射法直接检测底物或其类似物与酶形成的中间复合物（包括酶和底物）的相对位置。如溶菌酶是 129 个氨基酸残基组成的水解酶，其底物是细菌细胞壁的肽多糖。X 射线衍射图表明，溶菌酶是一个近似椭圆形的分子，其表面有一条与分子长轴平行的沟，其长度恰巧容纳底物分子中的 6 个糖单位，当底物与酶结合时，糖单位与沟壁活性中心的基团如 Glu35，Asp52 等以氢键相连，其中第 4 个糖环形成半椅式构型后与酶更紧密的结合，酶的催化作用是水解第 4 个糖环中的糖苷键水解。

（2）化学法：酶催化作用的特点之一是具有高度专一性，即一种酶只能与一种或一类底物结合并催化底物转化为产物。有的物质虽与底物类似并能与酶的活性中心结合，但不能发生催化反应，产生稳定的酶－底物复合物。根据这些原理，可采用同位素标记的化学修饰法或假底物法使底物与酶结合形成复合物；然后将该复合物蛋白水解，即可找出带有同位素的氨基酸，即为必需的基团。如乙酰乙酸脱羧酶的底物是乙酰乙酸，在反应条件下，二者结合并脱去羧基后，加入还原剂硼氢化钠即可形成稳定的酶－底物复合物，水解后即可发现赖氨酸的 ε－NH_2 为酶的必需基团。再如采用对某些化学基团有特异结合作用的化学修饰剂与酶分子活性中心中的基团共价结合，导致酶活力丧失，也可推知修饰后使酶失活的化学基团。化学方法的缺点是只能修饰极性基团，无法区别被修饰的羧基来自 Asp 或 Glu，氨基是 ε－还是 α－末端氨基，也难以得知活性基团的定位（氨基酸序号）。

（3）蛋白质工程：是近年来发展起来的研究酶必需基团与活性中心的最先进方法。

其原理是将酶相应的互补 DNA（cDNA）定点突变，此突变的 cDNA 表达出只有一个或几个氨基酸置换的酶蛋白，再测定其活性就可以知道被置换的氨基酸是否为活力所必需。该法具有化学修饰法无法比拟的优点，目前在国际上已被普遍应用。

二、酶催化作用的机制

酶促反应的特点之一是具有极高的催化效率，酶的催化效率若以等摩尔与一般催化剂比较，可高出 $10^7 \sim 10^{13}$ 倍。酶为什么有如此高的催化效率？研究其作用机制对人工制造高效的模拟酶和改造酶无疑可提供重要的理论基础。

（一）酶－底物复合物的形成与诱导契合假说

在酶促反应中，酶（E）首先与底物（S）以非共价键结合，形成 ES 复合物，然后发挥催化作用，释出产物（P）和酶。

1890 年 Emil Fischer 提出了酶－底物复合物的锁钥学说（lock－key theory），认为酶的活性中心与底物分子在结构上具有紧密互补的关系，如同锁和钥匙的关系一样。此学说可以解释某些酶的专一性，但不能解释酶催化反应的可逆过程。20 世纪 60 年代的 Daniel Koshland 提出了诱导契合假说（induced－fit hypotheory），认为酶与底物相互接近时，其结构相互诱导、相互变形和相互适应，进而相互结合（见图 5－2）。

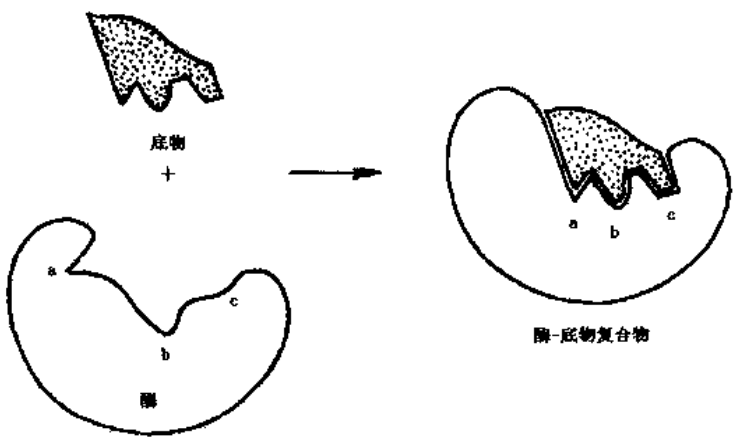


图 5－2 酶与底物结合的诱导契合假说示意图

在化学反应体系中，反应开始时，处于“基态（ground state）”的分子占绝大多数。这些分子的能级较低，也较稳定，不能形成产物，只有在吸收一定的活化能后，使基态分子变成“过渡”态（transition state）或活化态，才能转变成产物。

在供能不变的情况下，反应所需的活化能越高即能障越大，反应速度就越低。当酶与底物接近时，酶的构象改变有利于与底物结合；底物在酶的诱导下发生构象改变，使其接近于过渡态。ES 的形成又能稳定底物的过渡态，使底物进入过渡态所需要的活化能较无酶存在时大为降低，反应速度也就显著加速（见图 5－3）。

（二）酶催化反应机制的学说

1. 趋近效应与定向效应　当酶催化两个以上底物进行反应时，诸底物均结合到酶的活性中心，使它们相互接近，这就是趋近效应（approximation）。趋近效应可使底物在活性中心区的有效浓度提高数千至数万倍，大大增加了底物间碰撞的机会，使反应速度极大的提高。

酶与底物结合后才能起催化反应，所以仅靠近还不行，还需定向，即底物与酶的活性中心结合时，可诱导酶的构象变化，使底物和酶的活性中心互补契合，并使底物的反应基团彼此严格定向，使酶与底物的结合达到最有利形态，使反应加速。这种现象称为定向效应（orientation）。

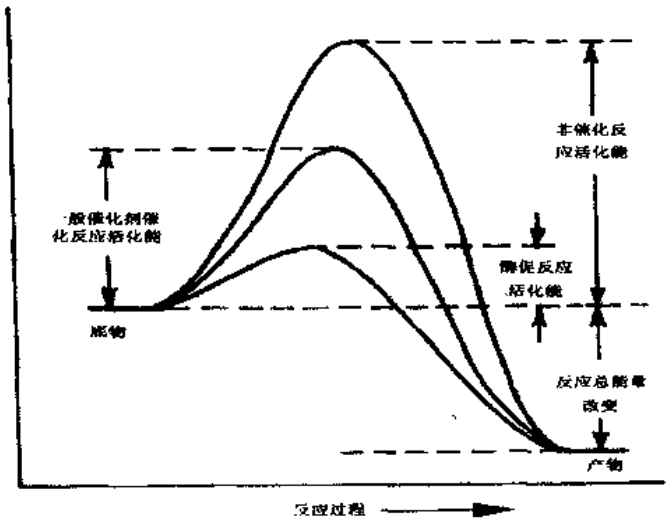
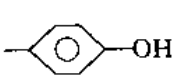
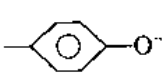
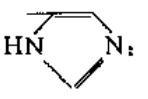


图 5—3　酶促反应活化能的改变

2. 张力效应　酶与底物结合，酶活性中心中的某些基团或金属离子可以改变底物敏感键的电子云分布，产生“电子张力”而易于断裂。也可以使底物的构象改变，接近于过渡态而易于反应。这种现象称为张力（strain）效应。

3. 共价催化理论　包括亲核催化和亲电子催化两类。因酶分子活性中心的一些基团如羟基、巯基和咪唑基，分别带有多电子的原子如 O，S 和 N，可以提供电子去“攻击”底物上相对带正电（亲电子）的原子，这种现象称为亲核攻击（nucleophilic attak），—OH，—SH 和咪唑氮等能进行亲核攻击的基团称为亲核基团（nucleophilic group）。亲核攻击通过形成酰化酶共价中间物，减少活化能而加速化学反应。有些酶分子中具有能与电子亲和的原子中心，能从底物得到一对电子形成酶—底物共价化合物，继而催化底物转变成产物，这种现象称为亲电子催化（electrophilic catalysis）。

表 5—1　酶活性中心中广义的酸碱基团

广义的酸性基团（质子供体）	广义的碱性基团（质子受体）	该基团的 pKa
$-COOH$	$-COO^-$	3.96（Asp），4.32（Glu）
$-NH_3^+$	$-NH_2$	10.80
$-NH-C(=NH_2^+)-NH_2$	$-NH-C(=NH)-NH_2$	12.48
$-SH$	$-S^-$	8.33
		10.11
		6.00

4. 酸碱催化 酸碱催化有两种主要类型，即狭义酸碱催化和广义酸碱催化。前者是指在催化反应体系中，缓冲溶液浓度不变时，化学反应速度随 pH 值的变化而改变；后者是指在催化反应体系中，pH 值不变，反应速度随缓冲溶液浓度的改变而变化。生物体内接近中性环境，H^+ 或 OH^- 的浓度仅为 10^{-7} mol/L 左右，且其浓度不可能有明显变化，狭义酸碱催化难以进行。然而，酶活性中心的某些化学基团有的是质子的供体（酸），有的是质子的受体（碱），可以发挥广义酸碱催化作用。表 5－1 列出了常见的广义酸碱功能基团。

三、酶促反应动力学参数

酶促反应动力学（kinetics of enzymic reaction）是研究酶促反应的速度行为及其影响因素。其目的在于了解机体正常和异常代谢，此外尚可了解酶促反应过程的本质。影响酶促反应速度的因素甚多，如温度、pH、抑制剂、激动剂等。但在其他条件固定时，反应速度决定于酶的浓度和底物的浓度。在酶促反应动力学研究中，最常用到的参数是米氏常数（K_m）和最大速度（V_{max}）。

（一）米曼氏方程式

1902 年 Henri 在研究蔗糖酶水解蔗糖的反应中发现，在底物［S］浓度很低时，随着底物浓度的增加，反应速度（V）急剧上升，二者成正比关系。随着底物浓度的进一步增高，反应速度不再成正比例增加。以 V 对［S］作图是一条矩形双曲线（见图 5－4），提示酶有饱和效应；即底物达到一定浓度时，可以使酶与底物的结合达到饱和。根据这一发现，Henri 提出了酶促催化反应的中间产物学说，即：

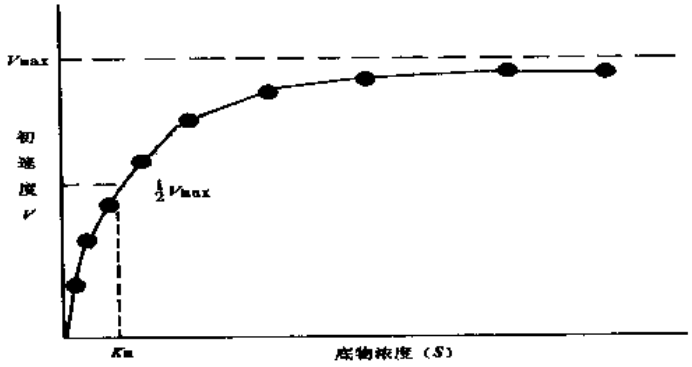


图 5－4 底物浓度对酶促反应速度的影响

$$E + S \underset{K_2}{\overset{K_1}{\rightleftharpoons}} ES \xrightarrow{K_3} E + P$$

酶 底物 中间产物 酶 产物

1913 年 L. Michaelis 和 M. L. Menten 根据中间产物学说提出了反应速度与底物浓度关系的数学方程式，即米—曼氏方程式，简称米氏方程式（Michaelis equation）：

$$V=\frac{V_{max}[S]}{K_m+[S]}$$

式中 V_{max} 为最大反应速度（maximum velocity），［S］为底物浓度，K_m 为米氏常数（Michaelis constant），V 是在不同［S］时的反应速度。当底物浓度很低（［S］$\ll K_m$）时，反应速度与底物浓度成正比。当底物浓度很高（［S］$\gg K_m$）时，$V \cong V_{max}$，反应速度达最大速度，再增加底物浓度也不再影响反应速度。

米氏方程式的推导基于如下假设：①测定的反应速度为初速度，即反应刚刚开始，此时，产物的生成是极少，逆反应可以不考虑；②在初速度时，[S] 远大于 [E]，[S] 的变化可忽略不计。

（二）K_m 与 V_{max} 的意义

K_m 和 V_{max} 是酶促反应动力学最常用到的参数，对二者的意义概述如下：

1. K_m 值等于酶促反应速度为最大速度一半时的底物浓度。K_m 能反应酶与底物的亲和力，二者呈反比关系。

2. K_m 值是酶的特征常数之一，只与酶的结构、酶所催化的底物和反应条件（如温度、pH、离子强度）等有关，与酶的浓度无关。对于同一底物，不同的酶有不同的 K_m 值。

3. V_{max} 是酶完全被底物饱和时的反应速度，与酶的浓度呈正比。

4. 根据酶的 K_m 值，可计算某一底物浓度时的反应速度与 V_{max} 的比值，即被底物饱和的分数。也可以估计反应速度接近于 V_{max} 时的底物浓度，一般在 $10\sim100K_m$ 之间。

（三）K_m 和 V_{max} 值的测定

通过改变底物浓度测得酶反应速度，作图可得出 K_m 值和 V_{max} 值。作图法测定 K_m 和 V_{max} 的方法有多种，最常用的是双倒数作图法（double reciprocal plot）又称林—贝氏（Lineweaver—Burk）作图法。该方法是将米氏方程式等号两侧取倒数，得到双倒数方程式：

$$\frac{1}{V}=\frac{K_m}{V_m}\cdot\frac{1}{[S]}+\frac{1}{V_m}$$

若以 $1/V$ 对 $1/[S]$ 作图，可得一直线（见图 5－5），其纵轴上的截距为 $1/V_{max}$，横轴上的截距为 $-1/K_m$。此作图法除用于求取 K_m 值和 V_{max} 值外，还可用以判断可逆性抑制反应的性质。竞争性抑制反应中，K_m 增大，V_{max} 不变；非竞争性抑制时，K_m 不变，V_{max} 降低；反竞争性抑制时，K_m 减少，V_{max} 降低。

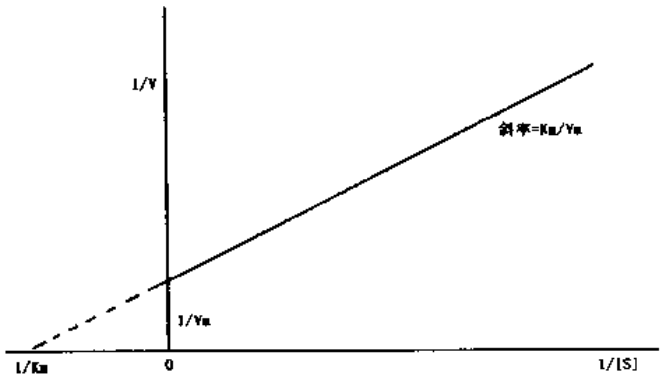


图 5－5　双倒数作图法

酶促反应速度还受温度和 pH 等因素的影响，只有在最适温度（optimum temperature）和最适 pH（optimum pH）条件下，酶促反应速度最大。故在测定 Km 值和 Vmax 值时应在最适温度最适条件下进行。

第二节 蛋白酶工程

由于酶在工业、农业和医药学上的广泛应用以及生物样品中酶来源的限制，人们期待用人工方法制造具有高效而专一催化作用的酶。又由于酶尚存在不稳定性和专一性不理想等缺点，酶学家也希望用现代科学手段来改造酶，蛋白质工程的建立，则为酶蛋白一级结构的改造提供了先进的方法。本节重点介绍几种蛋白酶改造的基本原理及应用。

一、枯草杆菌蛋白酶的定点改造

（一）结构及生物学特性

枯草杆菌蛋白酶（subtilisin）是由枯草芽孢杆菌（*Bacillus subtilis*）合成的一种丝氨酸蛋白酶。枯草杆菌蛋白酶依其来源不同分为 Carlsberg，BPN′，E，BL，YaB 等多种。大部分枯草杆菌蛋白酶是由 270 多个氨基酸残基组成。BPN′与 E 的区别是 43 个氨基酸残基发生替换，Carlsberg 与 E 的区别是 89 个氨基酸残基替换，以及第 56 位上的残基缺失，BL 与 BPN′的区别是在 275 个氨基酸残基上有 103 个残基替换再加上 160—163 位的 4 个残基缺失。南极嗜冷芽孢杆菌分泌的碱性蛋白酶原由 419 个氨基酸残基组成，成熟的酶蛋白由 309 个氨基酸残基组成，与嗜温枯草杆菌蛋白酶有 52％的氨基酸序列相同。

枯草杆菌蛋白酶催化蛋白质水解为氨基酸，在有机溶剂中也能催化多肽的合成。这些碱性蛋白酶具有重要应用价值，被广泛应用于洗涤剂、制革及丝绸工业。该酶的三维结构及编码基因都已基本搞清楚，而且有合适的克隆、表达系统以及准确的定量分析酶的催化活性的方法，因此它是进行蛋白质工程研究的合适对象。它的许多理化性质，诸如酶催化速率、底物专一性、pH 效应以及抗氧化作用、耐热和耐碱等性能，都已经应用基因定点诱变技术作了改进。

（二）定点改造

迄今为止对枯草杆菌蛋白酶所做的定点诱变已有 400 多个，在理论上和应用上都取得了相当的进展和成功，以下举例作一简略介绍。

1. 提高酶的稳定性　枯草杆菌蛋白酶已被发展为商用的纺织品洗涤剂，但实践中发现早期生产的枯草杆菌蛋白酶的洗涤剂无法与漂白剂共同使用。因为漂白粉剂会使枯草杆菌蛋白酶失活。生物化学分析揭示，造成这种失活效应的分子本质是，枯草杆菌蛋白酶的第 222 位甲硫氨酸残基（Met－222）发生了氧化作用，它可使该酶的活性丧失 90％。为了解决这一问题，科学工作者应用定点诱变技术，每次用一种非野生型的氨基酸密码取代 Met－222 的密码子，获得一系列相应的突变基因。然后将这些突变基因分别克隆到表达载体上，结果表达出 19 种不同的枯草杆菌蛋白酶突变体。蛋白酶活性检测显示，以 Cys－222 取代 Met－222，其活性超过野生型的。遗憾的是它同样会被漂白剂所氧化。若以 Ala－222 取代 Met－222，虽然其活性仅为野生型的 53％，但其抗氧

化性大大提高，在 1mol/L H_2O_2 中保存 1 小时，仍然保留原有活性，与漂白剂共同使用时，其活性不会被抑制。因此，含有经过这种定点改造的枯草杆菌蛋白酶的洗涤剂，是可以同漂白剂联合使用的。

酶的化学本质是蛋白质，其热稳定性比较差，枯草杆菌蛋白酶也不例外。要提高该酶在工业和商业领域的应用价值，提高酶的热稳定性具有重要意义。利用定点诱变技术改造枯草杆菌蛋白酶，以提高其热稳定性的研究工作比较多，并已获得了令人鼓舞的成果。如将 Ser－128、Leu－104 突变为 Gly－128 和 Val－104 的 Savinase 分别使突变酶热稳定性提高了 10.4℃。利用交错延伸技术（staggered extension process，STEP）将两个编码热稳定性枯草杆菌蛋白酶 E 的基因重组，野生酶在 65℃半衰期为 5 分钟，由于 PCR 错配获得 5 种突变酶，在 65℃半衰期是野生酶的 3－8 倍，而利用 STEP 技术重组得到 5 株突变株，获得 3 种突变酶，在 65℃是野生酶的 25－50 倍。将南极嗜冷芽孢杆菌所产的枯草杆菌蛋白酶通过定位诱变，以 Asp－85 取代 Thr－85 修饰钙离子配位体可将酶的热稳定性提高至嗜中温枯草杆菌蛋白酶的程度。这一突变酶的活性提高了 2 倍，并使得适应冷的突变酶的总的活性是嗜中温枯草杆菌蛋白酶的 20 倍。

2. 改变底物特异性与酶的活性　蛋白质工程的一个主要目的是开发出具有新的活性和改变了底物特异性的酶。枯草杆菌蛋白酶有较广的底物特异性，通过它的两个特异性口袋 S_1 和 S_4，与底物的侧链氨基酸残基 P_1 和 P_4 反应。通过改变 S_1 口袋的电荷、疏水性或立体结构已成功的提高了底物残基 P_1 和 P_2 的特异性。近年来通过对 S_4 口袋的氨基酸替换也成功的改变了对底物残基 P_4 的特异性。如将枯草杆菌蛋白酶 BL 的 Val－104 突变成 Tyr－104，突变酶催化四肽的活性大大提高，但对酪蛋白的水解仅提高了 26%。枯草杆菌蛋白酶 YaB 的三维结构显示，Gly－124 和 Gly－151 位于 S_1 口袋的腰部的两边，通过定位诱变获得了不同的 Gly－124 和 Gly－151 的突变酶，与催化酪蛋白和肌原纤维蛋白相比，它们相对提高了对弹性蛋白－苔红素的催化活性。

3. 提高酶在有机相中的反应效率　有机酶催化是目前一个颇具基础性和应用性的研究领域，酶在非水相溶剂中具有独特的性质和功能。如枯草杆菌蛋白酶在有机相中能催化多肽的合成，而且比常规的方法具有多方面的优越性。但是，枯草杆菌蛋白酶在有机溶剂中的半寿期缩短，其活性比在水相中低得多。为了克服这一缺点，已有人对枯草杆菌蛋白酶在有机相中的结构和机理进行了研究，而且有人已经应用基因定点诱变技术，对该酶的生理生化性能做了多方面的修饰改造，增添了许多新的特性。包括①将表面电荷的数量减少到最少的程度，以便降低融合作用的能量。如将枯草杆菌蛋白酶 8397 的表面电荷去除，即 Lys－43 替换为 Asn，Lys－256 替换为 Tyr，Asp－181 替换为 Asn，提高了酶在有机相中的稳定性；②增加内部极性及疏水作用性能。如将 BPN′ 的 Gly－166 替换成 Asn，催化活性提高 178 倍。这可能是极性突变引起的过渡态的稳定性改变造成的。动力学研究也显示，在 S_1 结合位点的极性突变能提高酶的催化活性；③导人构象限制用以降低蛋白质变性倾向。现在已分离到了大量的、溶剂稳定性得到显著提高的枯草杆菌蛋白酶突变体。

（三）应　用

枯草杆菌蛋白酶对于玷污服装的各种蛋白质，有广泛的特异性消化能力。因此，枯

草杆菌蛋白酶早在几十年前就被开发成商用的纺织品洗涤剂，目前仍然是最基本、最重要，也是使用最广泛、最成熟的洗涤用酶。国外 20 世纪 60 年代推出枯草杆菌蛋白酶，目前洗涤剂的 90%为加酶洗涤剂。国内 1979 年推出枯草杆菌蛋白酶，1990 年国内枯草杆菌蛋白酶产量约 13000 吨，加酶洗衣粉占洗涤总量的 10%。另一方面枯草杆菌蛋白酶作为蛋白质工程的研究对象，对它的研究不仅仅改善了其性能，更为重要的是丰富了蛋白质工程的理论，为其他酶，尤其是其他工业酶的性能改造提供了范例和理论指导。

二、组织型纤溶酶原激活剂的定点改造

(一) 结构及生物学特性

组织型纤溶酶原激活剂（tissue plasminogen activator，tPA），是一种丝氨酸蛋白酶，人体各种组织均可合成。它能够切割纤溶酶原形成纤溶酶，从而激活纤溶作用(fibrinolysis)，使血栓溶解。因此 tPA 是治疗栓塞性疾病的理想药物。然而由于天然 tPA 的半寿期仅有 5 分钟，因此在治疗中难以持续的发挥药效作用。通过基因定点诱变技术，在体外对组织型纤溶酶原激活剂的氨基酸残基进行改造，以延长其半寿期，即可有效地提高临床应用价值。

人 tPA 基因位于第 8 对染色体 $p^{12}-q^{11.2}$，全长 36594bp，由 14 个外显子和 13 个内含子组成。14 个外显子分别编码 tPA 的各个功能区。从人类细胞培养液中提纯或用重组技术制备的 tPA 为单链（sctPA）糖蛋白，分子量为 6.8 万～7.2 万，含 527 个氨基酸残基（见图 5－6），很容易被纤溶酶、组织激肽释放酶、因子 X 将 275Arg～276Ile 肽键裂解，使 tPA 转化为由重链和轻链组成的双链活化物（tctPA），重链（MW30882）位于单链 tPA 的 N 末端，轻链（MW28126）位于 C 末端，两者通过一个二硫键连接。轻链含有 His－325、Asp－374 和 Ser－481 组成的酶活性中心（P 区），重链按顺序分为指状结构区（finger domain，F 区）、EGF 区（G 区）、Krigle1 区（K_1）和 Krigle2 区（K_2），其中 EGF 区位于第 44 至第 92 残基之间，与人的 EGF 有同源性，通过与肝脏 tPA 受体结合而介导 tPA 从血清中的清除。K_1 和 K_2 区位于第 93 至第 297 残基之间，呈环状结构。这些区域可能是

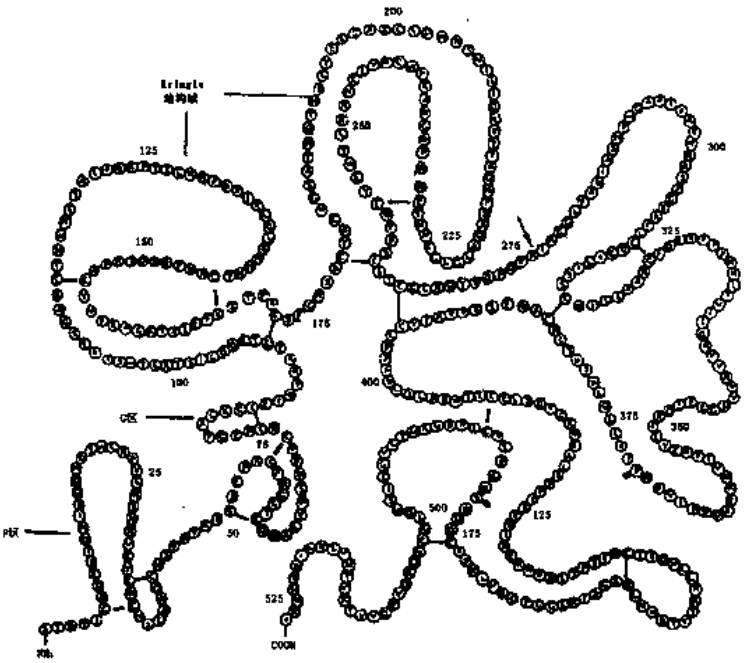


图 5－6 tPA 的分子结构

独立发挥生物学作用的功能域。

tPA 的等电点在 7.8～8.6 之间，最大组分等电点在 8.2；其 pH 在 5.8～8.0 时较为稳定，最适 pH 为 7.4。另外，tPA 在枸橼酸抗凝血浆中不稳定，其活性丧失速度与温度有关。正常人血中 tPA 水平约为 25μg/L，半衰期 45 分钟。tPA 的主要作用是特异性裂解纤维蛋白溶酶原中的 Arg561－Val－562 肽键。使纤溶酶原激活。同时它通过 F 区和与纤维蛋白结合成 tPA 纤维蛋白复合物能高效特异性的激活血凝块中的纤溶酶原，产生具有胰酶专一性的广谱蛋白酶即纤溶酶，使纤维蛋白降解成可溶性产物，达到血块溶解，栓塞的血管重新畅通的目的。

（二）定点改造

近年来，研究者从研究 tPA 的结构入手，以构效关系为原则，利用 DNA 重组及基因定点诱变技术构建了一系列 tPA 变异体，它们在延长其体内的半衰期、增加对纤维蛋白亲和力及提高催化活性等方面均有较大改善，较之天然 tPA 有更广阔的应用前景。

1. 将 tPA 分子的 3 个糖基化位点进行改造。选择性的将这些位点的 Glu 突变为 Arg，得到一种突变体 tPA，其在体内的半衰期比野生型 tPA 延长 20 倍，纤溶活性也有所提高。

2. 将 tPA 的 N 端 F 和 G 区编码序列的核苷酸做缺失处理，表达的 tPA 突变体的半衰期为野生型 tPA 的 25 倍。

3. 已知 tPA 的 PAI－1 结合位点是^{296}Lys－His－Arg－Arg－Ser－Pro－^{302}Gly 及^{304}Arg。利用基因定位诱变及 DNA 重组技术构建了一种 tPA 缺失突变体 tPA－del，这个突变体缺失了 tPA 的 K_1区和^{296}Lys－^{302}Gly 的 7 个氨基酸残基，由此获得了对 PAI－1 的抗性，在大鼠体内的半衰期延长约 5 倍。

此外，为了获得临床治疗效果更好的基因工程产品，科学家还尝试进行 tPA 的嵌合体表达。嵌合体表达是选择性地使 tPA 的不同功能区联络，通过优先组合来提高其特性，减少其副作用。

（三）应　用

tPA 是一种高效特异性溶血栓药物，在临床上已广泛应用的另外 2 种溶血栓药物链激酶（streptokinase，SK）和尿激酶（urokinase，UK）相比具有显著的优点：①血栓选择性高，只对纤维蛋白有特异亲和性，对纤维蛋白原亲和力很低，由其激活形成的纤溶酶，溶栓活力比尿激酶多 5～10 倍，但几乎不激活循环血液中的纤溶系统，可直接静脉注射；②对中毒性休克引起的弥散性血管内凝血疗效显著；③tPA 是存在于人体内的天然蛋白质，不像链激酶那样会引起变态反应，是一种理想的溶栓剂，一般用于治疗心肌梗死、异体肾移植后排斥反应、肾病综合征、下腔静脉血栓形成等疾病。

此外，现已发现血中 tPA 含量变化与多种疾病密切相关，可作为某些疾病临床诊断指标：①纤溶系统功能的判断指标；②诊断血栓形成的指标；③凝血疾病预后诊断的指标；④肝功能评价指标。

第三节 葡萄糖异构酶工程

葡萄糖异构酶（glucose isomerase，GI）又称D—木糖异构酶（D--xylose isomerase），1957年首次发现于嗜水假单胞菌中，随后有近百种放线菌被鉴定为产GI的菌株。1967年美国首次成功地进行了GI工业化应用，至今西方主要的淀粉加工业都已转向GI技术。目前国际公认，GI是研究蛋白质结构与功能关系、建立完整的蛋白质工程技术最好的模型之一。

一、结构特点及生物学性质

（一）生物学性质

GI能催化D—葡萄糖至D—果糖的异构反应，是工业上从淀粉制备高果糖浆（high fructose corn syrup，HFCS）的关键酶，而且可将木聚糖异构化为木酮糖，再经微生物发酵生产乙醇。GI的最适pH通常在7.0～9.0之间，在偏酸性条件下，大多数种属的GI活力很低。GI最适温度一般在70～80℃，这决定于缓冲液、底物浓度、激活剂、稳定剂及反应时间等条件。不同菌株产生的GI的热稳定性不同。如乳酸杆菌和埃希杆菌GI的热稳定性较差，链霉菌和枯草芽孢杆菌GI在高温下相当稳定，而嗜热高温菌GI的热稳定性最高。GI的活性及稳定性还受金属离子的影响。Mg^{2+}，Co^{2+}及Mn^{2+}等对该酶有激活作用，而Ca^{2+}，Hg^{2+}及Cu^{2+}等则起抑制作用。

（二）结构特点

不同来源GI序列不尽相同，GI的序列同源性在相近种属中较高。如密苏里游动放线菌（Actinoplanes missourinesis，AM）和小瓶属菌株GI的序列同源性达93.4%，而大肠杆菌和枯草杆菌GI与链霉菌的同源性仅为24.2%和26.3%。在已知GI的序列中，除了活性部位的14个氨基酸全保留外，还有其他25个氨基酸也相同。

不同来源的GI在晶体中均以紧密的四聚体形式存在，单体分子量为40～50Ku。每个单体分两个结构域，N端的主结构域由8股α/β桶（α/β barrel）构成，它包含两个同轴圆柱，内圆柱由8条平行的β折叠片构成，外圆柱由8股与β折叠片交替相邻的α螺旋构成，α螺旋的肽链走向与β折叠片成反平行。C端的小结构域由几段α螺旋无规卷曲成一个远离N端的环状结构，参与单体间的相互作用及活性部位的构成。

GI的每一个单体均有一个口袋状活性中心，位于平行β桶的近C端口部。每一个活性中心包含M_1和M_2两个二价金属离子结合位点。M_1是结构离子，与Glu－180，Glu－216，Asp－244及Asp－286的羧基氧原子，His－219的咪唑基团及一个溶剂分子的氧原子配位成键，M_2在异构化过程中位置偏离原处0.1nm左右。Glu－216是与M_1，M_2配位的桥梁。在上述氨基酸构成的极性区域外，有一个由Phe－93，Trp－15，Trp－136和phe－25′形成的“高度疏水背景”。

二、定点改造葡萄糖异构酶

（一）提高热稳定性

7 号淀粉酶链霉菌 M_{1033} 菌株所产葡萄糖异构酶（Streptomyces diastaticus No. 7 strain M_{1033} glucose isomerase，SM33GI）是工业上大规模以淀粉制备高果糖浆的关键酶，但用于工业生产，其热稳定性还不够理想。蛋白质结构模拟显示，Pro 有较少的构型自由度，其吡咯烷环可使 β 转折或无规卷曲结构更稳定，能增加蛋白质空间结构的刚性，提高蛋白质的热稳定性。朱国萍等学者通过分子结构模拟，设计将 Gly138 改为 Pro，以提高 SM33GI 的热稳定性。他们用寡核苷酸定点诱变方法对 SM33GI 的基因进行体外定点突变，将 Gly－138 替换成 Pro，构建了 GI 突变体 G138P。含突变体的重组质粒 pTKD－GIG138 在 E. coli K_{38} 菌株中表达获得突变型 GIG138P，与野生型 GI 比较实验表明：①GIG138P 的热失活半衰期约是野生型的 2 倍；②GIG138P 的最适反应温度提高了 10～12℃；③GIG138P 的比活性与野生型 GI 相当。图 5－7（a）为 SM33GI 晶体结构中 Gly138 附近氨基酸空间分布的立体图，图 5－7（b）为模拟 Pro 取代 Gly138 之后的情况。由图可见，将 Gly138 替换成 Pro 后，Pro 的吡咯烷环侧链刚好能够充填，Gly138 由于无侧链基团而留下的空洞，使得整个分子结构更加紧密、更加稳定。

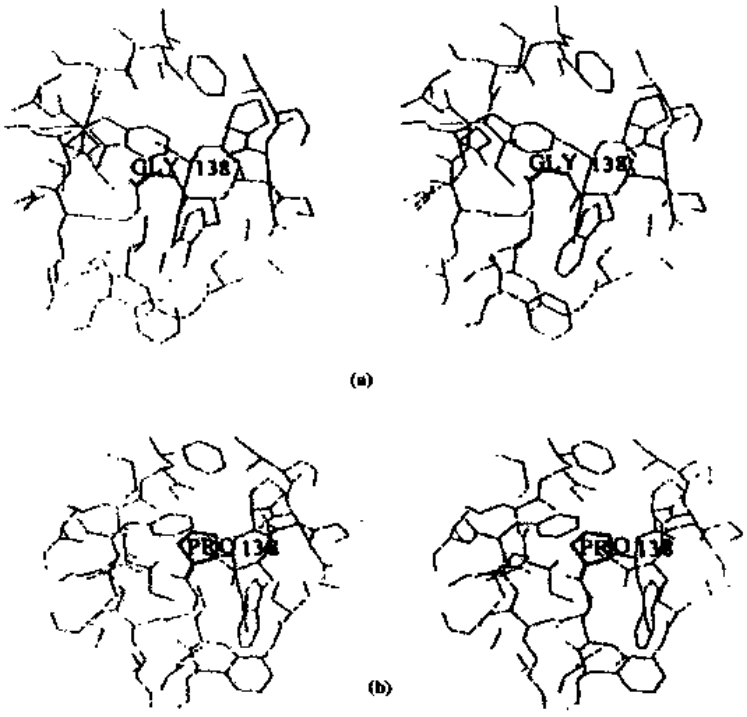


图 5－7　野生型 GI 和 GIG138P 立体结构示意图

Mrabet 构建的密苏里游动放线菌 GI 突变体酶 K253R，其半衰期从野生型的 670h 延长到 1550h。三点突变体 K309R/K319R/K323R 在 60℃时的半衰期是野生型的 3 倍。最稳定的突变体 G70S/A73S/G74T 则是通过加强活性二聚体界面的相互作用而改善了热稳定性。

（二）降低最适 pH 值

大规模由淀粉制备高果糖浆生产中，降低反应 pH 可增加果糖得率。但是野生型 GI 最适 pH 较高，故降低 GI 最适 pH 对大规模生产高果糖浆具有重要意义。利用定点诱变技术改造 GI 基因已获得多种突变株，使突变 GI 的最适 pH 降低。如锈赤链霉菌的 D56N 和 E221A 突变酶体，当 pH 从 8.5～8.8 降至 6～7.5 时酶活性提高 40%，再如 AMGI 突变体 E186Q，其最适

pH 比野生型下降 1 个单位，且在 Mn^{2+} 存在条件下酶活性提高 2 倍。下面对 AMGI 突变体 E186Q 的构建及其意义作一简略介绍。

GI 活性中心部位有两个与二价金属离子结合的位点，当二价金属离子与之结合后可影响酶的催化活性。Mg^{2+}，Co^{2+} 及 Mn^{2+} 不仅能激活 GI，而且能影响对底物的选择性。如 AM 菌的 GI 和 Co^{2+} 结合时对木糖的活性最高，而与 Mn^{2+} 结合时对葡萄糖活性最高。酶分子中酸性氨基酸残基富含阳离子结合位点，其 pKa 使得在低于中性 pH 时有利于阳离子的结合。而野生型 GI 最适 pH 是 7.5，不利于阳离子结合。为了解决这一矛盾，Tilbeurgh 利用基因定点诱变技术，将 Glu－186 替换成 Gln－186，获得 E186Q 突变株，结果使 GI 的最适 pH 由 7.5 降至 6.25，在有 Mn^{2+} 存在时酶活性提高 2 倍。之所以选择 Glu－186 为定点诱变的靶位，是因为 Glu－186 是一个保守氨基酸，其靠近金属结合位点，但不直接参与酶与金属离子的结合，也不直接参与酶的催化活性。X 光衍射分析证明 Glu－186 的羧基第二个金属结合部位仅有 7Å，其解离所带负电荷足以影响金属离子与酶的结合。而以 Gln 替代 Glu，可降低其 pKa，并增强了在低 pH 时对金属离子的亲和性。

（三）改变底物专一性，提高催化效率

嗜热产硫梳状芽孢杆菌 GI 的定点诱变体 W139F，其 Km 降低，Kcat 升高。双突变体 W139F/V186T 及 W139F/V186S 的催化效率则分别高于野生型酶 5 倍和 2 倍。W139 是位于底物结合处的氨基酸残基，这说明底物特异性可以通过提高活性部位的底物结合口袋对葡萄糖的氢结合能力而得以改变。

SM33GI 双突变酶 G247D/G138P 的酶活性比野生型提高 45%。这可能因为 D247 附近活性残基的侧链均伸向活性中心，而 D247 的侧链向外伸展，而且负电性很强，改变了活性部位的电荷传递过程，而使酶活性提高。

三、应　　用

在大规模从淀粉制备果糖浆（HFCS）的过程中，葡萄糖转化率为 50%，而工业要求果糖含量高于 50%，提高反应温度及降低反应 pH 可增加果糖得率，在 HFCS 生产，玉米淀粉、α－淀粉酶及脱枝酶的液化、糖化和异构酶化在同一 pH 下进行，约为 4.5～5.0，因此酸、热稳定的 GI 能减少高温下有色碳酰副产物的生成，提高 HFCS 生产效益。由此可见，对 GI 进行改造，提高其酸、热稳定性，降低其最适 pH 值，是增加果糖得率，提高 HFCS 生产效益的关键所在。而蛋白质工程技术的发展，使 GI 的改造成为可能。相信很快即会获得性质改善的理想的 GI，使生产成本降低，工艺流程简化。科学家预计通过蛋白质工程改造的 GI 将成为最重要的工业用酶之一。

第四节　酶蛋白的定向进化

利用定点诱变的方法对天然酶蛋白质进行改造已有很多成功的例子。然而，该方法仅适用于三维结构以及结构与功能的相互关系基本清楚的蛋白质，目前我们对这类关系

的认识还很肤浅。对于那些结构与功能关系尚不清楚的酶蛋白，定点突变是无能为力的。近年来有人提出“定向进化”的观点，为酶的结构与功能的研究开辟了崭新的途径。

一、酶的定向进化的原理与思路

（一）定向进化的原理

酶的体外定向进化（directed evolution of enzyme in vitro）是改造酶蛋白质分子的一种新策略。它不需事先了解酶的定向结构和催化机制，在实验室模拟自然进化机制，通过由易错 PCR、致突变菌株诱变等方法对编码酶蛋白质的基因进行随机诱变，由 DNA 改组（DNA shuffling）、随机引发重组和交错延伸等方法进行突变基因体外重组，设计高通量筛选方法来选出需要的突变株。它不仅可快速生产工业上有用的新酶，而且为研究蛋白质的结构与功能的关系开辟了崭新的途径，极大地拓展了蛋白质工程学的研究和应用范围。

酶的体外定向进化的基本原理是，在待进化酶基因的 PCR 扩增反应中，利用 Taq DNA 多聚酶不具有 $3'\rightarrow5'$ 校对功能的性质，配合适当条件，如降低一种 dNTP 的量等，以很低的比率向目的基因中随机引入突变，构建突变库，凭借定向的选择方法，选出所需性质的优化酶（或蛋白质），从而排除其他突变体。简言之，定向进化就是随机突变加选择，前者是人为引发的，后者虽相当于环境，但只作用于突变后的分子群，起着选择某一方向的进化而排除其他方向突变的作用，整个进化过程是在人为控制下进行的。

（二）定向进化的基本思路

定向进化是在不了解蛋白质的结构信息情况下，根据人为设定目标，制备具有特定性质和功能的特殊蛋白质。在制定定向进行研究方案时需要考虑如下几个关键问题：①随机诱变方法是通过引进随机的碱基替换，进而筛选理想的突变体。一般有益突变的频率很低，绝大多数突变是有害的，当突变频率太多时，几乎无法筛选到有益突变体；但突变频率也不能太低，否则未发生任何突变的野生型将占据突变群体的优势，也很难筛选到理想的突变体。一般认为，理想的突变频率为每个目的基因的碱基替换在 1.5 到 5 个左右。②组成一种蛋白质分子的氨基酸经排列组合可产生巨大的顺序空间，也即可获得数量极大排列方式，其中绝大多数排列方式没有反应功能，尤其是我们想要的功能。因此在考虑实验方案时，最好是选择一个性状最接近人们期望的酶分子作为起点；③要从大量的突变体库中获得理想的个体，最重要的问题是筛选方法，筛选方法越好，成功的可能性越大。因此，应建立有效且灵敏的选择方法，确保检测出由单一氨基酸取代而引起的功能变化。

二、定向进化的常用技术

（一）易错 PCR

易错 PCR（error－prone PCR）是指在扩增目的基因的同时引入碱基错配，导致目的基因随机突变，改变 PCR 的条件，通常降低一种 dNTP 的量（降至 5%～10%），或加入 dITP 来代替减少的 dNTP，即会使 PCR 易于出错，达到随机突变的目的。然而，经一次突变的基因很难获得满意的结果，由此发展出连续易错 PCR（sequential error－prone PCR）策略。即将一次 PCR 扩增获得的有用突变基因作为下一次 PCR 扩增的模板，连续反复地进行随机诱变，使每一次获得的小突变累积而产生重要的有益突变。

Chen 等人利用此策略定向进化枯草杆菌蛋白酶 E 的活性获得成功，所得突变体 PC3 在高浓度二甲基酰胺（DMF）中，酶催化效率 Kcat 是野生酶的 256 倍。将 PC3 再进行两个循环的定向进化，产生的突变体 13M 的催化效率比 PC3 高 3 倍，比野生酶高 471 倍。

用易错 PCR 法进行定向进化改造，关键在于突变率的控制。另外，该法较为费力、耗时，一般适用于较小的基因片段（<800bp）。

（二）DNA 改组

DNA 改组（DNA shuffling）是 Stemmer 于 1994 年建立的模仿自然进化的一种 DNA 体外随机突变方法。所谓 DNA 改组，就是将 DNA 拆散后重排，即将一种基因或具有结构同源型的几种基因在 DNase I 的作用下随机酶切成小片段，这些小片段之间有部分的碱基序列重叠，它们通过自身引导 PCR（self－priming PCR）延伸，并重新组装成全长的基因，这一过程被称为再组装 PCR（reassembly PCR）。DNA 改组的基本过程包括四个步骤（见图 5－8）。①目的基因的准备：根据需要选择一个基因或其片段，也可以是几个序列上具有较高同源性的基因；②DNase Ⅰ酶切：将目的基因随机切割成约 10～50bp 或 300bp 左右的小片段；③不加引物的 PCR：在 Taq 酶的作用下将 DNase Ⅰ切割后的 DNA 重叠小片段重新连接起来，在此过程中可能发生许多突变和重组；④加入引物的 PCR：加入目的基因片段两端的引物，使连接好的 DNA 得到扩增，筛选正突变。得到正突变子又可以重复进行改组，使性状

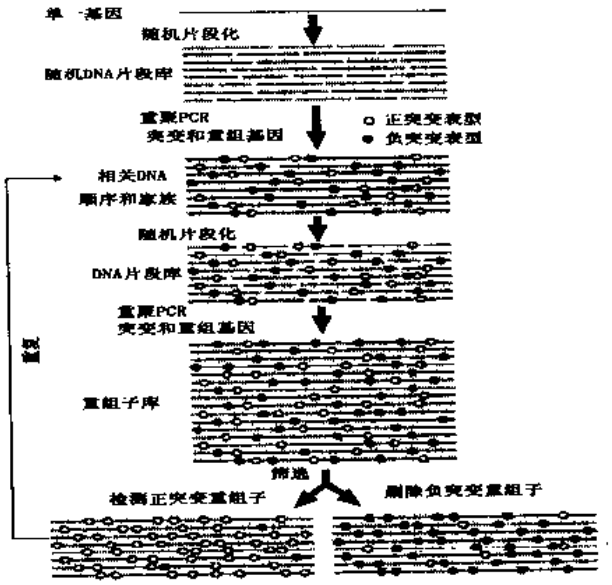


图 5－8 DNA 改组原理

进一步提高。

DNA 改组的目的是创造将目的基因群中的突变尽可能组合的机会，导致更大的变异，最终获取最佳突变组合酶。在理论和实际上，它都优于重复寡核苷酸引导的诱变和连续易错 PCR。通过 DNA 改组，不仅可加速积累有益突变，而且可使酶的 2 个或更多的已优化性状合为一体。

（三）随机引发重组

随机引发重组（random－priming recombination，RPR）是 Aronld 于 1998 年首先报道。其基本原理是以单链 DNA 为模板，配合一套随机序列引物，先产生大量互补于模板不同位点的 DNA 小片段，由于碱基的错误掺入和错误引发，在随后的 PCR 反应中，它们互为引物进行合成，伴随重组，再组装成完整的基因（见图 5－9），克隆到表达载体上，随后筛选。该法较 DNA 改组有以下优点：①RPR 直接利用单链 DNA 或 mRNA 为模板；②在 DNA 改组中，片段重新组装前必须彻底除去 DNase I，而 RPR 方法省去了 DNase I 切割成片段过程，因而使基因的重组组装更容易；③合成的随机引物具有同样长度，无顺序倾向性，保证了点突变和交换在全长的后代基因中的随机性；④随机引发的 DNA 合成不受 DNA 模板长度的限制。

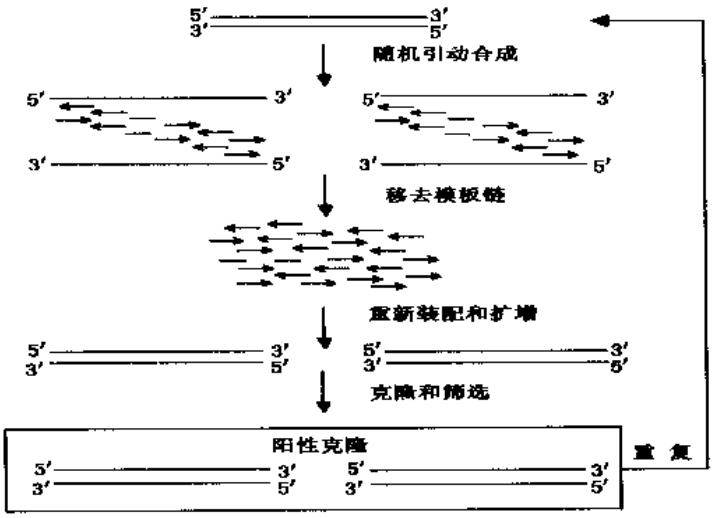


图 5－9　体外随机引发重组原理

（四）交错延伸技术

交错延伸（stagger extension process，StEP）是 Aronld 等人于 1998 年建立的一种新的体外重组方法。其原理是，在 PCR 反应中把常规的退火和延伸合并为一步，并缩短其反应时间（55℃，5s），从而只能合成出非常短的新生链，经变性的新生链再作为引物与反应体系内同时存在的不同模板退火并延伸。此过程重复进行，直到产生完整的基因长度，结果产生间隔的与不同模板序列互补的新生 DNA 分子（见图 5－10）。该技术已成功地重组了由易错 PCR 产生的 5 个热稳定的枯草杆菌蛋白酶 E 的突变体，得到了热稳定性进一步提高的重组酶。

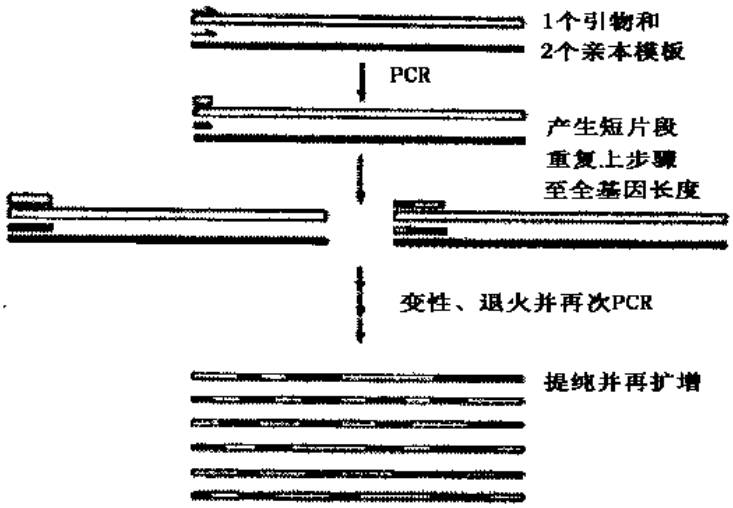


图 5－10　交错延伸原理

交错延伸法重组在单一试管中进行，不需分离亲本 DNA 和产生的重组 DNA。它采用的是变换模板机制，这正是逆转录病毒所采用的进化过程。该法简便有效，为酶的体外定向进化提供了又一强有力工具。

（刘贤锡）

第六章　抗体工程

自从 Kohler 和 Milstein 于 1975 年提出用杂交瘤技术成功研制出单克隆抗体以来，尝试将该技术应用于肿瘤和疾病治疗的研究从没有间断过。由于单克隆抗体对相关抗原有高度特异性，作为诊断试剂在 20 世纪 80 年代取得了重大进展，然而在体内的应用，由于单克隆抗体是异源性蛋白质，易引起人抗小鼠抗体，影响了其疗效，因此在小鼠杂交瘤技术出现不久，就开始了人杂交瘤抗体技术的研究，但由于单克隆抗体应用于肿瘤治疗的研究没有取得实质性突破，而使抗体研究进入低谷。进入 90 年代以来，抗体作为治疗性试剂的研究再度引起了人们的高度关注，并取得了令人鼓舞的成果。肿瘤抗体导向基因治疗的研究呈现回头的趋势，特别是有关抗体工程和抗体分子遗传学的研究有了进一步发展。应用基因工程改造抗体产生了基因工程抗体、人源化抗体、小分子抗体和胞内抗体等具有广泛应用前景的新型抗体。抗体库和噬菌体展示系统技术的创建，被称为抗体技术中又一次革命性进展。抗体库技术用细菌克隆取代了 B 细胞克隆表达抗体，不经细胞融合，甚至不经免疫制备针对任何抗原的抗体，为人抗体的制备提供了一条十分有效的途径。为了研究抗体的理化性质、分子结构与功能，以及应用抗体于临床疾病的诊断、治疗及预防都需要人工制备抗体。目前，根据制备的原理和方法可分为多克隆抗体、单克隆抗体及基因工程抗体三类。基因工程抗体和抗体库技术的进展已经使抗体的制备和应用进入了一个全新的时代。

第一节　概　　述

抗体是一个极不均一的分子群，有不同的类和亚类，又表现出各不相同的抗原特异性，但不同的抗体有着相同的结构，即都是由两条相同的重链和两条相同的轻链组成的。重链（H）分子量为 50～70kD 的多肽链，轻链（L）较短，分子量为 23kD，每一条肽链可分为两个区域，重链氨基端的 1/4 或 1/5 以及轻链氨基端的 1/2 在不同的抗体分子间变化较大，称为可变区（V），其余部分称为恒定区（C）。抗体分子用木瓜蛋白酶消化以后可以产生两个相同的 Fab 段和一个 Fc 段，前者为抗原结合部位，后者与抗体的效应功能相关。Fv 段为重链和轻链的可变区相互作用形成的抗原结合部位，抗体可变区的氨基酸变异并非均匀分布，而是集中在几个较小的部位，这些区域被称为超变区或者互补决定区（complementarity－determining region，CDR）。轻链和重链分别有

三个 CDR 区域。抗体的轻链可分为 κ 链和 λ 链，分子量皆为 23kD，哺乳动物抗体重链可分为 μ，α，γ，δ，ε 链，在人类 γ 链又可分为四个亚类：γ1，γ2，γ3，γ4，α 链可分为 α1 和 α2，μ 为 IgM，α1 为 IgA1，α2 为 IgA 2，γ1 为 IgG1，γ2 为 IgG2，γ3 为 IgG3，γ4 为 IgG4，δ 为 IgD，ε 为 IgE，重链的类和亚类决定了抗体的类和亚类。

一、多克隆抗体

大多数抗原是由大分子蛋白质组成，但只是抗原上有限部位的特殊分子结构能与其相应抗体结合，此部位称为抗原决定簇（antigenic determinant）或表位（epitope）。一种天然抗原性物质（如细菌或其分泌的外毒素以及各种组织成分等）往往具有多种不同的抗原决定簇，而每一决定簇都可刺激机体的一种抗体形成细胞产生一种特异性抗体。在机体内存在许多种抗体形成细胞（即 B 细胞），每种抗体形成细胞只识别其相应的抗原决定簇，当受抗原刺激后可增殖分化为一种细胞群，这种由单一细胞增殖形成的细胞群体可称之为细胞克隆（clone）。同一克隆的细胞可合成和分泌在理化性质、分子结构、遗传标记以及生物学特性等方面都完全相同的均一性抗体，亦可称之为单克隆抗体。在早期，传统的抗体制备方法是将一种天然抗原经各种途径免疫动物，由于抗原性物质具有多种抗原决定簇，故可刺激产生多种抗体形成细胞克隆，合成和分泌抗各种决定簇抗体分泌到血清或体液中，故在其血清中实际上是含多种抗体的混合物，这种用体内免疫法所获得的免疫血清称为多克隆抗体，也是第一代抗体。由于这种抗体是不均一的，无论是对抗体分子结构与功能的研究或是临床应用都受到很大限制，因此如何获得均一性抗体成为关注的问题。

二、单克隆抗体

体内免疫法很难获得单克隆抗体（monoclonal antibody，McAb）。如能将所需要的抗体形成细胞选出并能在体外进行培养即可获得已知特异的单克隆抗体。1975 年德国学者 Kohler 和英国学者 Milstein 将小鼠骨髓瘤细胞和经绵羊红细胞（sheep red blood cell），SRBC）免疫的小鼠脾细胞在体外进行两种细胞融合，结果发现部分形成的杂交细胞既能继续在体外培养条件下生长繁殖又能分泌抗 SRBC 抗体，称这种杂交细胞系为杂交瘤（hybridoma）。这种杂交瘤细胞既具有骨髓瘤细胞能大量无限生长繁殖的特性，又具有抗体形成细胞合成和分泌抗体的能力。它们是由识别一种抗原决定簇的细胞克隆所产生的均一性抗体，故称之为单克隆抗体。应用杂交瘤技术可获得能引起小鼠抗体应答的所有抗原的单克隆抗体。这种用杂交瘤技术制备的单克隆抗体可视为第二代抗体。单克隆抗体由于纯度高、特异性强、可以提高各种血清学方法检测抗原的敏感性及特异性，但单克隆抗体多为双价抗体，与抗原结合不易交联为大分子集团，故不易出现沉淀反应。单克隆抗体的应用大大提高了对各种传染病和恶性肿瘤诊断的准确性。单克隆抗体亦可与核素、各种毒素（如白喉外毒素或蓖麻毒素）或药物通过化学偶联或基因重组制备成导向药物（targetting drug）用于肿瘤的治疗，是一种新型免疫治疗方法，有可能提高对肿瘤的疗效。单克隆抗体亦可用于对各种免疫细胞及其他组织细胞表面分子的检测，这对免疫细胞的分离、鉴定、分类及研究各种膜表面分子的结构与功能都具有重

要意义。

三、基因工程抗体

自 1975 年单克隆抗体杂交瘤技术问世以来，单克隆抗体在医学中被广泛应用于疾病的诊断及治疗。因为目前绝大多数单克隆抗体是鼠源的，临床重复给药时体内产生抗鼠抗体，使临床疗效减弱或消失。所以，临床应用理想的单克隆抗体应是人源的，但人－人杂交瘤技术目前尚未突破，即使研制成功，也还存在人－人杂交瘤体外传代不稳定，抗体亲和力低及产量不高等问题。目前较好的解决办法是研制基因工程抗体（genetically engineering antibody），以代替鼠源单克隆抗体。基因工程抗体兴起于 20 世纪 80 年代早期。这一技术是将对 Ig 基因结构与功能的了解与 DNA 重组技术相结合，根据研究者的意图在基因水平对 Ig 分子进行切割、拼接或修饰，甚至是人工全合成后导入受体细胞表达，产生新型抗体，也称为第三代抗体。基因工程抗体包括嵌合抗体、重构抗体、单链抗体、单区抗体及抗体库等。其中以嵌合抗体研究的较多，也较成熟。单链抗体及单区抗体虽具有结构简单、分子小等优点但其临床应用的前景尚待证实。

第二节　抗体及其基因结构与抗体多样性的分子基础

机体内有成千上万的 B 细胞，每一个 B 细胞克隆都表达有特异性的抗原受体，具有不同的特异性，说明每一个个体都存在着数量巨大的抗体，即抗体的多样性，它是由免疫球蛋白可变区，尤其是 CDR 区决定的。另一方面，免疫球蛋白的恒定区具有类、亚类、型和亚型等有限变化，恒定区的变化与可变区的多样性完全分离，同一个亚类重链可具有多种不同的可变区，而同一个可变区又可与不同类或亚类的恒定区组成重链。自从五十年代以后，随着免疫球蛋白氨基酸序列资料的积累使免疫球蛋白分子的基本结构得到阐明，就提出了两个最基本的问题：从可变区和恒定区的结构和功能特点来看，似有着不同的遗传来源，它们如何会存在于同一条肽链？如此巨大数量的可变区是如何形成的？1965 年 Dryer 和 Benner 提出了一个大胆的设想：免疫球蛋白肽链是由两组基因编码，一组数量巨大，含有可变区的遗传信息，另一组小得多，含有恒定区的信息，这与当时公认的“一个肽链由一个基因编码”的原则相矛盾。按照这个设想，一条肽链可能由两个基因片段编码，它还意味着可变区和恒定区基因必须拼接在一起才能转录并翻译成蛋白质分子。由于当时实验技术手段不足，支持 Dryer 和 Bennet 假设的实验证据直到 10 年以后才出现。现在我们已知道 Ig 肽链由不止一个基因编码的设想的原则是正确的，但免疫球蛋白的基因结构和重组远较 Dryer 和 Bennet 设想的要复杂得多。

一、抗体基因的基本结构

免疫球蛋白的 κ 链基因、λ 链基因和重链基因位于不同的染色体。在生殖细胞系或在除 B 细胞以外的其他细胞中，为可变区和恒定区编码的基因片段在基因组 DNA 上是分开的，轻链和重链的基因结构又互不相同。

（一）轻链基因的基本结构

1. κ链 人和鼠的κ链基因结构非常类似，由三种分离的基因片段组成，即 V_κ，J_κ和 C_κ，C_κ编码恒定区（见图 6－1），J 基因段编码κ链可变区第 96～108 位氨基酸，人和鼠都有五个 J 基因段，小鼠的 $J_\kappa 3$ 没有功能，因其 3’ 端缺乏 RNA 拼接信号（内含子 5’ 端的剪接信号为 GT，但在小鼠 $J_\kappa 3$ 的 3’ 端是 CT）。V_κ编码κ链可变区的 1～95 位氨基酸，其确切数目还不确定，目前估计约 100 多个 V_κ基因，其中有相当一部分不能合成有功能的蛋白质，称做假基因（pseudogene）。出现假基因的原因有，编码区出现终止密码子、因插入或缺失造成的阅读框架位移等。假基因虽不能直接与 J 拼接成有活性的可变区基因，但它们仍可作为可变区信息库通过基因转换（gene conversion）或不对称重组增加κ链的多样性。

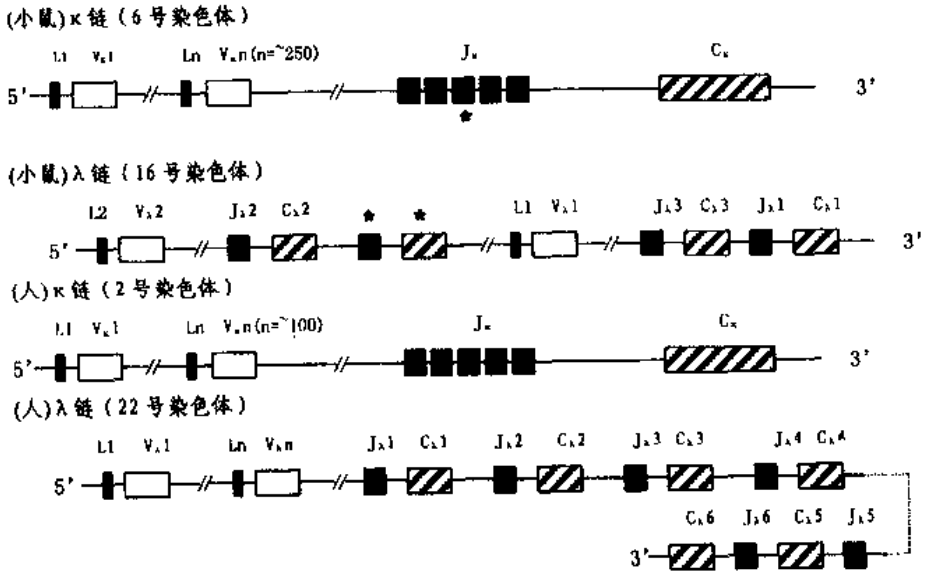


图 6－1 抗体轻链基因结构（星号表示无功能的假基因）

典型的 V_κ基因含有上游调控序列，前导序列，轻链蛋白的编码序列及 3′端的重组信号序列。在前导序列内含有一个内含子。前导序列所编码的前导肽对抗体分子进入粗面内质网腔，最终到达细胞膜表面或分泌到胞外有重要作用。当免疫球蛋白的肽链进入粗面内质网腔后，前导肽被信号肽酶切去，故前导肽不存在于成熟免疫球蛋白分子中。

V_κ基因可根据其序列相似程度不同分成不同的亚群，一般将 DNA 序列 80％以上相同的 V_κ基因片段视为一个亚群。在 DNA 杂交实验中，同一亚群的 V_κ基因可在较严格的条件下杂交，而不同的亚群之间的杂交信号很弱。亚群的划分是人为的，有些 V_κ基因难于归入某个亚群。但亚群在一定程度上反映了可变区基因在种系进化中的关系。目前在人类发现有七个亚群。在κ链基因位点中，同一亚群的 V_κ基因有分布同一区域的趋势，但不同亚群的 V_κ基因群常常混合分布，可能与进化过程中的基因转位（transposition）和基因扩增有关。

2. λ链　λ链也由三个基因片段编码，V_λ和J_λ组成可变区，V_λ为1～95位氨基酸编码，J_λ为96～108位氨基酸编码，恒定区由独立的外显子编码。由于λ链有不同的亚型，每一个亚型有各自的恒定区基因。λ链基因的结构与κ链有所不同，每一个C_λ有其自己的J基因片段，形成J－C结构，而J_κ基因集中在一起形成J基因群。人类有七个C_λ基因，$J_\lambda1$与V_λ基因相距14kb，在七个C_λ基因中$C_\lambda1$相应于mcg^+亚型，$C_\lambda2$相应于ke^-Oz^-型。人类V_λ基因目前鉴定出十个亚群，最近已克隆出了36个V_λ基因，其中12个为假基因，从DNA印迹杂交实验估计人类V_λ基因总数可能在70个左右。在小鼠则仅有两个V_λ基因，这与小鼠血液中只有5%Ig含有λ轻链相一致。小鼠有四个$J_\lambda-C_\lambda$基因，每个V_λ基因跟随着两个$J_\lambda-C_\lambda$，$V_\lambda1$后面是$J_\lambda1-C_\lambda1$和$J_\lambda3-C_\lambda3$，$V_\lambda2$后面是$J_\lambda2-C_\lambda2$和$J_\lambda4-C_\lambda4$，其中$J_\lambda4-C_\lambda4$是假基因，在J_λ和C_λ之间有一个1.2～1.4kb的内含子。可看出，$V_\lambda2$可以和任何一个J－C重组，而$V_\lambda1$只可以和$J_\lambda3$，$J_\lambda1$发生重组。

（二）重链基因的基本结构

重链基因和轻链基因的结构有两点主要区别：一是可变区由三个基因片段编码，它们是V_H（variable），D（diversity）和J_H（joining），二是重链所有同种型的恒定区基因集中在一起，位于可变区基因段的下游（见图6－2）。

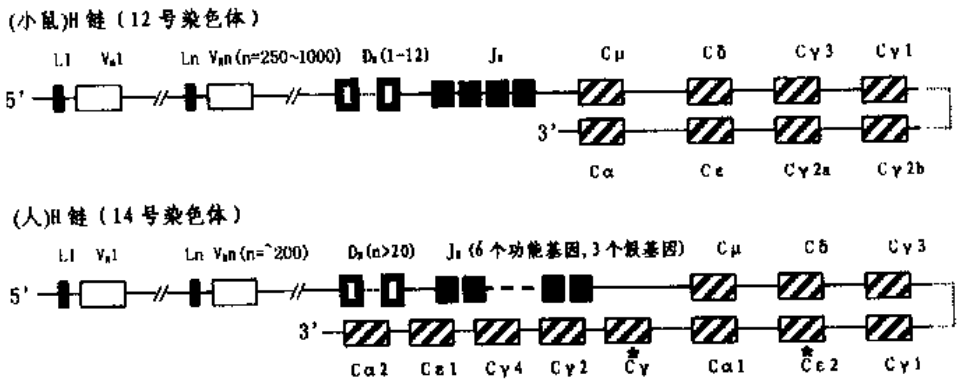


图6－2　抗体重链基因结构（星号表示无功能的假基因）

V_H基因与轻链的V基因类似，由两个外显子组成，第一个外显子为大部分前导肽编码，第二个外显子为前导肽羧基端四个氨基酸和可变区1～95位氨基酸残基编码，两个外显子之间有大约100个碱基对左右的内含子。人类V_H的数目现已比较清楚，大约有100个左右具有功能的V_H基因，共分六个亚群。另有约几十个假基因，具体数目尚不清楚。假基因可通过间接方式增加抗体的多样性。小鼠V_H数目尚不确定，估计数百左右，共分十个亚群。

D基因片段所编码的肽段构成了CDR3的主要成分。它没有固定的阅读框架，且长短不一，短者可编码三个氨基酸，长者可编码十几个氨基酸，因此其确切数目难以估计。D基因群在V_H基因群和J_H基因群之间，分布在大约70～80kb的DNA范围内，在

V_H基因之间也可以发现散在的D基因片段。

在D基因群的下游是J_H基因群，J_H为重链可变区羧基端16～21处氨基酸编码，其氨基端的4～9个氨基酸残基与D片段共同组成CDR3。人类有九个J基因段，其中有3个假基因，小鼠有4个J基因。距J_H基因群下游约7000个碱基为重链恒定区基因群。

重链恒定区基因群含有所有重链同种型的恒定区基因，在小鼠有八个恒定区基因，分布在200kb的范围内，在人类有11个恒定区基因，其中有两个假基因，其确切长度尚不清楚。恒定区基因之间除C_μ和C_δ之间相隔4～5kb外其他都相距数万个碱基，每一个恒定区基因由4到6个外显子组成，每一个功能区及铰链区由独立的外显子编码，3′端的1～2个外显子为膜型Ig的穿膜区和胞内区编码。

二、可变区基因的重组及其机制

由于可变区分成数个基因片段，每一个基因片段以基因群的方式存在，因此只有通过基因重组使各基因片段，即V—J或V—D—J连接起来才能形成有功能的完整的可变区基因。

(一) 轻链可变区基因的重组

轻链可变区基因由V—J重组完成（见图6—3），任何一个V_κ或V_λ基因与某个J_κ或J_λ经基因重组连在一起，即可形成一个有功能的轻链可变区基因。在V—J重组时，接头处可发生核苷酸错位和变化，从而增加CDR3的多样性。

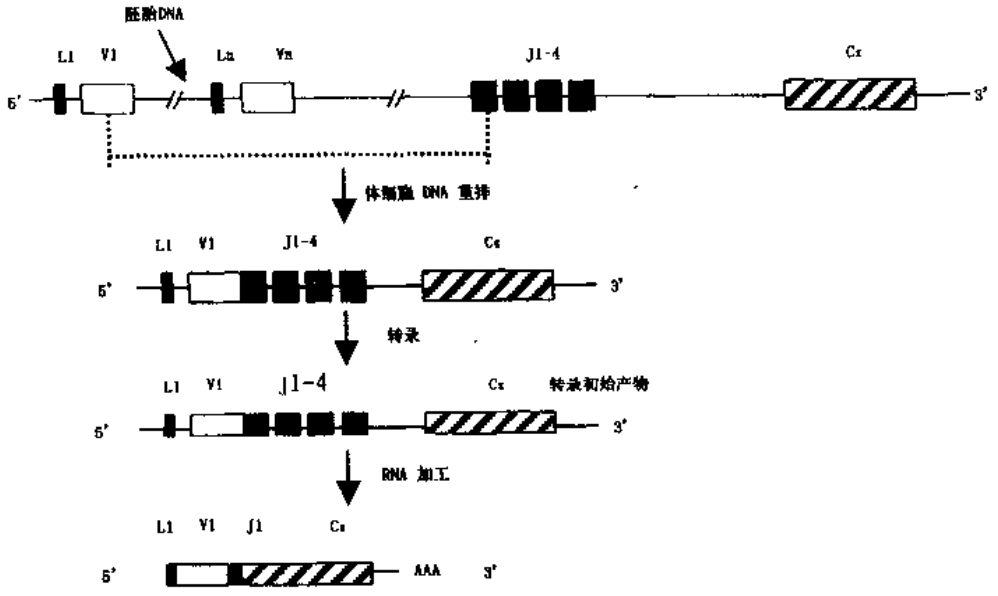


图6—3 Ig蛋白κ轻链基因重排及转录机制

(二) 重链可变区基因的重组

重链可变区基因的重组由于多了D基因片段而较轻链更为复杂。首先发生的是D

—J 重组，将 1 个 D 基因片段和 1 个 J_H基因连接起来，然后进行 V—DJ 重组，一个 V_H 基因与重排后的 DJ 连接形成有功能的重链可变区基因。D—J 和 V—DJ 重组时接头处的变化比轻链 V—J 接头处的变化要大得多，如在 D—J 和 V—DJ 重组时末端的核苷酸可以被酶解，还可通过末端核苷酸转移酶的作用加上新的核苷酸，称作 N 区（non-germline encoded nucleotides），它大大增加了重链可变区 CDR3 的多样性。

（三）可变区基因重组的机制

1. 重组信号序列　V—（D）—J 重组的机制还未完全阐明，但通过对这些基因片段两侧 DNA 序列的分析，在 V_H和 V_L基因 3′端，J 基因的 5′端和 D 基因的两端发现了高度保守的重组信号序列（recombination signal sequence，RSS），它包括一个七核苷酸组成的七聚体（heptamer）和一个九个核苷酸组成的九聚体（nonamer），V_L和 J_L，V_H和 D_H以及 D_H和 J_H之间的七聚体和九聚体为反向互补序列（见图 6—4）。如 J 基因 5′端的七聚体序列为 CACTGTG，九聚体序列为 ACAAAAACC，分别与 V_L3′ 端的七聚体 CACAGTG 和九聚体 GGTTTTTGT 为反向互补序列，这些互补序列可以形成茎—襻状结构（stem and loop structure）。七聚体和九聚体之间间隔序列的长度是固定的，为 12±1 或 23±1 个核苷酸，它决定了重组的模式，只有间隔 12 个碱基对和间隔 23 个碱基对的重组信号序列之间发生重组，而其他间隔长度的重组信号序列之间不发生重组，这就是 12/23 碱基对规律（12/23 base pair rule），由于 12 个碱基和 23 个碱基分别相当于 DNA 双螺旋的一或二圈，叫做 1+2 圈规律（1+2 turn rule）。这一规则保证了重组以正确的方式进行，尤其在重链，有 D—J，V—DJ 之间发生重组，而不会产生 V—J 重组。



图 6—4　可变区基因重组信号序列示意图

2. 重组模式　通过重组信号序列进行重组的具体模式有以下三种：

（1）缺失模式（deletion model）：以 V_L—J_L重组为例，V_L3′端的重组信号序列和 J_L5′端的重组信号序列通过七聚体和九聚体形成茎—襻状结构，将 V_L和 J_L基因端带到了一起，然后通过重组酶（recombinase）的作用进行切除和连接，将 V_L和 J_L连在一起形成完整的轻链可变基因，而 V—J 之间的襻，则通过七聚体末端之间的连接形成一个环状 DNA 分子。

（2）倒转模式（inversion model）：当重组信号序列同处于两种基因段的同一侧时，

上述重组过程不是造成两个重组基因间 DNA 序列的环化和缺失，而是这段 DNA 的倒转，倒转后 V—J 连接成可变区基因。在 D 基因段两端均有重组信号序列，因此 D—J 重组可通过以上两种模式进行。在基因组中发现有反向排列的 V 基因说明这些 V 基因通过倒转模式进行重组，在体外模式实验中发现，在两个方向的重组信号序列都存在的情况下，倒转模式和删除模式发生的几率几乎相等。因此，根据基因组中 V_κ 基因群的排列方向，大约可有半数 V_κ—J_κ 的重组可能涉及倒转模式。

(3) 姊妹染色体交换模式 (sister chromatid exchange model)：通过姊妹染色体之间的不对称交换或重组 (unequal crossing over or recombination) 可在两条染色体之间进行 V—(D)—J 重组，这种机制在理论上可能存在，但尚未发现直接证据。

三、CDR3 的多样性

Ig 可变区中以 CDR3 的变异性最大，有人估计人重链 CDR3 的多样性可达 10^{14} 以上，事实上 CDR3 的多样性是抗体多样性的主要来源，V—(D)—J 重组的接头部位都位于 CDR3 内，重组过程中有多种机制可增加其多样性。

(一) 接头处多样性 (junctional diversity)

在 V—(D)—J 重组时重组信号序列末端的连接和编码序列末端的连接有很大差异，信号序列末端的连接一般较准确，未发现有碱基丢失的现象，添加新的核苷酸的现象也很少发生。但编码序列末端的连接一般不准确，常可见到 1～10 个核苷酸的丢失或添加，所添加的核苷酸有两种，一种叫做 N 区或 N 核苷酸 (N region，N nucleotides)，N 代表 non—germline encoded nucleotides，意为非生殖细胞系编码的核苷酸。这种新的核苷酸的添加是由末端脱氧核糖核酸转移酶 (TdT) 所催化的，一般以 GC 较为多见，N 区只发生在重链可变区的 DJ 和 VD 接头处。另一种叫做 P 核苷酸 (P nucleotides)，P 取自 Palindrom (回文结构)，因为它总是邻近于编码序列的末端并与编码序列末端的碱基互补。P 核苷酸一般只有 1～2 个碱基，只存在于完整的编码序列末端，因此人们推测 P 核苷酸的形成是 V—(D)—J 重组时的固有特性，它实际上来自 DNA 的互补链，通过两条链末端两个碱基间的相互连接，同时在互补链倒数第一、二或第二、三个碱基之间打开缺口，使该链末端的 1～2 个碱基墨守成规到另一条链上，造成突出的单链末端，再通过 DNA 聚合酶补平，形成 P 核苷酸。因此 P 核苷酸本质上是基因组所携带的信息。单链突出的末端也可成为 DNA 外切酶的底物被水解删除，造成编码区末端的缺失，并可通过 TdT 活性添加形成 N 区，因此在有 N 区形成时见不到 P 核苷酸。

(二) 来源于 D 基因段的多样性

由于重链可变区比轻链多一个 D 基因段，使重链的 CDR3 的多样性更为突出，除了基因组 DNA 多副本 D 基因段和上述接头处的变化外，还有其他机制可增加 D 片段的多样性，D 片段没有固定的阅读框架，随 V—D—J 重组时接头处的变化而变化。DIR 基因 (D_H genes with irregular recombination signals)：在 Ig 基因组 D 位点，除了常规

的 D 基因段，还发现有约 90 个碱基长的 DNA 序列，其两端有多个七聚体和九聚体重组信号序列，在七聚体和九聚体之间的间隔序列既有 12 个碱基也有 23 个碱基，因此它可以形成 V－D－DIR－J，V－DIR－D－J 及 V－D－DIR－D－J 等多种不同形式的重组，而增加 CDR3 的多样性。目前已发现有 7%的人类 CDR3 含有 DIR 基因序列。

（三）V－V（D）J 重组和 V（D）J－J 重组

V（D）J 重组完成以后的可变区基因还可以和上游的 V 基因再次重组，造成 V_H 替换（V replacement）或和下游的 J 基因再次重组，这些重组仍可产生接头处的变化，如形成 P 核苷酸及 N 区。关于其发生的机制目前认为存在于 V_H3' 端及重组后 VDJ 基因 3′端的与重组信号相类似的序列介导了 V－VDJ 和 VDJ－J 重组。

四、抗体多样性的来源

随着对 Ig 基因结构和重组机制了解的深入，抗体多样性产生的机制逐渐得以阐明。大体上可将抗体多样性的来源分成三部分：①胚系细胞所携带的遗传信息，它包括多副本的 V 基因，D 基因，J 基因及轻链重链等随机配对所形成的多样性，如按 50 个 V_H 基因，30 个 D 基因，6 个 J_H 基因，100 个 V_K 基因，5 个 J_K 基因计算，这样细胞固有的多样性可达：$50\times30\times6\times100\times5=4.5\times10^6$；②B 细胞个体发生过程中由于 V－（D）－J 重组时接头处的变化所产生的多样性，这种变化发生在 CDR3，包括重组时编码核苷酸的删除和添加，P 核苷酸和 N 区的形成等；③体细胞突变（somatic mutation）亦称高突变（hypermutation），发生于 B 细胞受到抗原刺激后的分化发育阶段。免疫系统在接触抗原以前通过各种机制已形成了数量巨大的免疫球蛋白群体，可以结合自然界中各种各样的抗原，但这还不足以对机体提供充分的免疫保护，因为大多数抗体的亲和力都比较低。抗原进入机体后刺激 B 细胞进一步分化，同时抗体基因发生体细胞突变，突变后亲和力得到改善的细胞克隆经抗原选择后具有较大的生长优势，从而在再次免疫反应中所产生的抗体具有较高的亲和力，这种亲和力增高的过程叫做亲和力成熟（affinity maturation）。

第三节　鼠单抗的人源化

单克隆抗体的问世极大促进了抗体在各个领域的应用，在临床检测中发展了许多以单抗为基础的检测手段，但单抗在体内治疗上的应用发展缓慢，其主要原因之一是人杂交瘤体系的低效性。迄今的单克隆抗体主要是鼠源性的，鼠源性抗体作为异种蛋白应用于人体可引起针对异种蛋白的免疫反应，产生抗小鼠抗体（human anti mouse antibody，HAMA），HAMA 既可影响单抗的治疗的效果，又可能诱发过敏反应。降低鼠单抗的异源性是解决这一难题的重要途径之一。杂交瘤－单克隆抗体技术和 DNA 重组技术所取得的成就促进了基因工程抗体的发展，为鼠单抗人源化的发展提供了必要的理论和技术。

一、鼠单抗恒定区的人源化

抗体分子是双功能分子，其与抗原特异性结合的活性存在于轻链和重链的可变区，即 Fv 段，抗体分子作为异种蛋白诱导免疫反应的抗原性则主要存在于恒定区，因此最初出现的人源化鼠单抗就是将鼠单抗的可变区与人抗体的恒定区拼接而成的嵌合抗体，称为人一鼠嵌合抗体（human anti mouse chimeric antibody）（见图 6－5）。它完整地保留了亲本鼠单抗的特异性和亲和力，同时大大降低了鼠单抗的异源性。

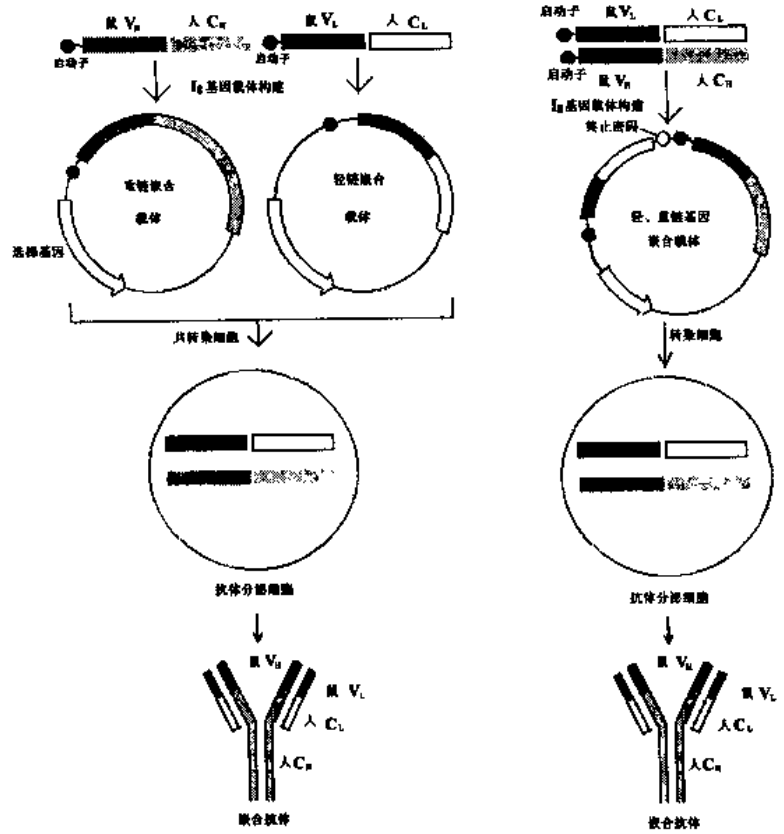


图 6－5　鼠单抗恒定区的人源化

二、鼠单抗可变区的人源化

虽然恒定区人源化部分消除了鼠单抗的异源性，但若彻底消除其抗原性，只能将可变区也进行人源化。可变区人源化有以下几种方法：

（一）人改型抗体

抗体的可变区有六个 CDR 区，直接与抗原接触，决定了抗体的特异性，其余部分只是作为支持 CDR 襻的支架，且结构极其保守，因此可将鼠单抗的 CDR 移植到人单抗的骨架区（见图 6－6），则可使人单抗获得鼠单抗的特异性，且极大地减少了鼠单抗的异源性。这种技术称为改型抗体（reshaped antibody），起初称为人源化抗体，但随着

抗体人源化技术的增加，容易引起混淆，因此可称为改型抗体。

（二）表面氨基酸残基的人源化

1991 年有人提出与 CDR 移植截然不同的人源化方法，其重要区别为：将鼠 Fv 段表面暴露的骨架区氨基酸残基中与人 Fv 不同者改为人源性，使 Fv 的表面人源化，消除其免疫原性而不影响 Fv 的整体空间构象，保留了抗原结合部位的结构。鼠单抗表面残基人源化后完全保留了亲本抗体的特异性和亲和力。

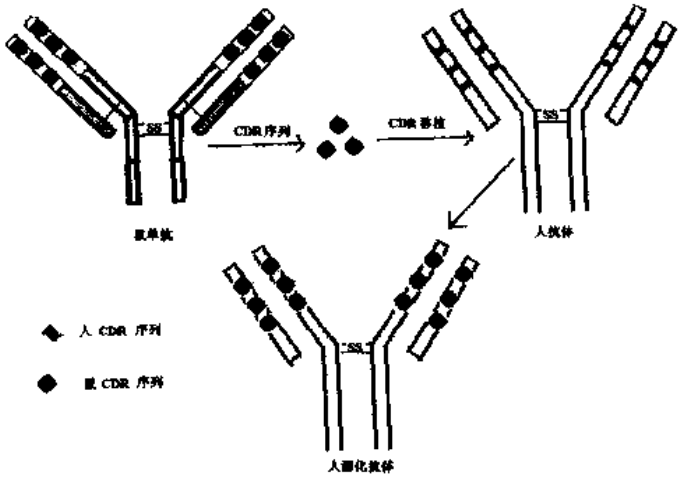


图 6－6　改型抗体

（三）表位印模选择

表位印模选择亦称为定向选择，是在噬菌体抗体库技术上发展起来的一种方法，其基本原理是：鼠抗体为模板，利用噬菌体表面展示系统，筛选出针对相同抗原决定簇的人抗体。将小鼠的一个可变区基因（重链或轻链）与人的另一个可变区基因文库（轻链或重链）配对构建成鼠－人抗体库，用对应的抗原筛选出有结合活性的克隆，得到人的轻链或重链基因；再将人的轻或重链基因与人的另一条链（轻链或重链）可变区基因文库组合成新的抗体库，再次以相应的抗原进行筛选，得到特异性与鼠单抗完全相同的人单抗。

第四节　小分子抗体

抗体分子的抗原结合部位局限在可变区组成的 Fv 段，从而可以构建分子量较小的具有抗原结合功能的分子片段，称为小分子抗体。由于小分子抗体的分子量较小而具有以下优点：①可以在大肠杆菌等原核细胞表达体系表达，通过细菌发酵生产，从而降低生产成本；②易子穿过血管壁或组织屏障，进入病灶部位，有利于对肿瘤等疾病的治疗；③不含有 Fc 段，不与 Fc 受体结合，可减少因广泛分布的 Fc 受体带来的不利影响；④在体内半衰期较短，这虽在导向治疗中是不利因素，但有利于体内毒性物质的清除；⑤易子进行基因工程改造，如构建抗体融合蛋白等。常见的小分子抗体包括：Fab 段，

Fv 及单链抗体（single—chain Fv，ScFv）等，分别介绍如下：

一、Fab 段

Fab 段由重链 Fd 和完整的轻链组成，二者通过一个二硫键连接，形成异二聚体，是完整抗体分子的三分之一，仅有一个抗原结合位点。早期通过木瓜蛋白酶对完整抗体分子酶解后分离纯化获得 Fab 段。从大肠杆菌表达有功能的抗体分子片段一度受阻，这是因为抗体分子与抗原的结合依赖子抗体分子的立体构象，即需要可变区正确的立体折叠、链内二硫键的形成，以及轻链重链可变区两个分子间互相作用形成正确的立体构象，这一过程在 B 细胞内是在粗面内质网完成，在大肠杆菌表达蛋白质常常需要在体外进行复性，这一过程对抗体这样复杂的分子效率很低。后发现大肠杆菌细胞壁的周质腔可提供类似于内质网的环境，将重链 Fd 基因与轻链基因 5′端接上细菌蛋白的信号序列（signal sequence），所表达的蛋白在细菌信号肽的引导下可分泌到周质腔（见图 6—7），信号肽被信号肽酶（signalpiptidase）切除，生成的 Fd 段和轻链在周质腔内完成立体折叠并通过二硫键形成异二聚体，成为有功能的 Fab 段。按照这一设想，在 1988 年从大肠杆菌成功地表达了有功能的 Fab 段，即分泌型表达。这一原理也可用于表达 Fv 段。

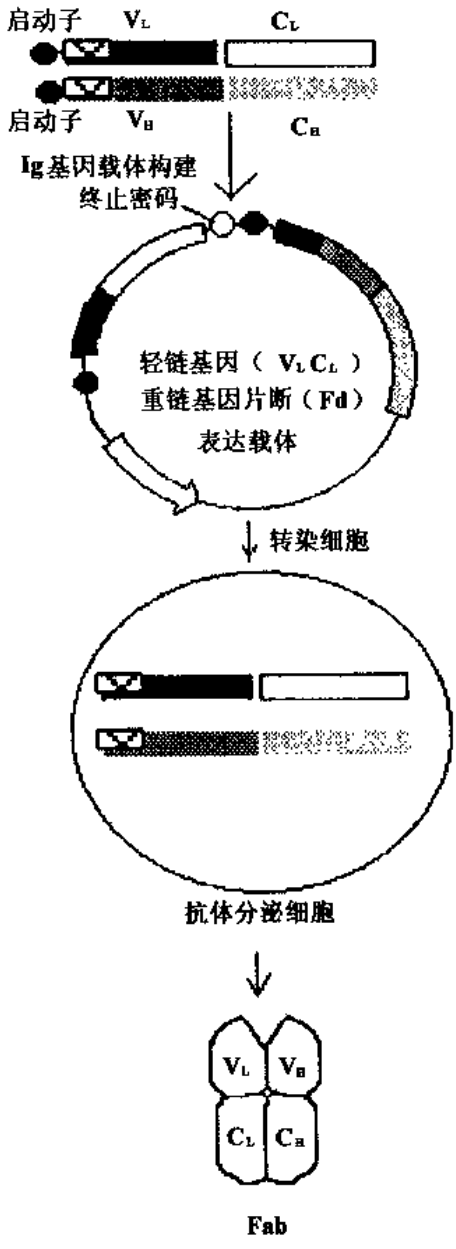


图 6—7 抗体 Fab 片段的表达（表示信号序列）

二、Fv 和单链抗体

Fv 段是抗体分子中保留抗原结合部位的最小功能片段，它由轻链可变区和重链可变区组成，二者通过非共价键结合在一起。用基因表达方法，通过将 V_H 和 V_L 转导到大肠杆菌周质腔，可获得有功能的 Fv 段。在 Fab 段，轻链稳定区和重链 CH1 之间有一个二硫键，使抗原结合部位的结构极为稳定，而在 Fv 段，重链和轻链可变区是由非共价键结合在一起，在浓度较低时有解离的倾向，因而极不稳定。不同 Fv 段的解离常数不同，一般在 10^{-5}～10^{-8}mol/L 之间，因此欲获得稳定的 Fv 段，需设法将两个可变区连接起来，目前有三种途径：一是用戊二醛处理 Fv 段，通过化学方法使 V_H 和 V_L 交联；二是通过对 Fv 段立体结构的分析选择适当的部位通过点突变引入半胱氨酸残基，使轻链和重链可变区

之间形成链间二硫键；第三种方法是在 DNA 水平将轻链和重链用一段适当的寡聚核苷酸（接头）连起来，使之表达成为一条单一的肽链，称为单链抗体（single chain Fv，ScFv），它更易于在原核细胞进行表达和在基因水平进行改造。这三种方法所获得的 Fv 段都基本上保持了 Fv 段的特异性和亲和力，而大大增强了稳定性。第一种方法比较繁琐，并且未经广泛的验证，采用的人很少。第二种方法使用的也不多，据报道其效果较好，较 ScFv 更好地保持了亲本抗体的亲和力，且更为稳定，但报导很少，其普遍实用性尚需进一步证实。ScFv 的制备则由于开展较早，方法较为成熟，而得到了较为广泛的应用。

（一）ScFv 的设计与构建

首先确定重链和轻链可变区基因，按照 Kabat 等的定义，κ 链可变区是从第 1 到第 108 位氨基酸，λ 链是 1～107 位氨基酸，V_H包括 1～113 位氨基酸，这是通过晶体结构分析确定的，V_H和 V_L的最后一个氨基酸残基涉及骨架之间氢键的形成。在设计 ScFv 时，有些人改动了可变区两端的氨基酸序列，在多数情况下未发现对抗体特异性和亲和力带来影响，但有时确可改变抗体的结合特性，因此在设计 ScFv 时应尽可能保持亲本抗体可变区序列的完整性。

ScFv 中的 V_H和 V_L是由接头（linker）连在一起，V_H和 V_L的取向可有两种方式，V_H－接头－V_L或 V_L－接头－V_H，两种构建方式都被证明不影响 ScFv 的特异性和亲和力，但有报道在大肠杆菌分泌型表达时，ScFv 的分泌效率与 V_H和 V_L的取向有关，尽管表达总量相同，V_L－接头－V_H比 V_H－接头－V_L的分泌量要高 20 倍。

ScFv 分子中接头的设计对保持亲本抗体的亲和力有重要影响，它应当不干扰 V_H和 V_L的立体折叠，并且不对抗原结合部位造成妨碍。为不使 Fv 的立体结构变形，接头的长度应不短于 3.5nm。由于相邻肽键的距离约为 0.38nm，接头应至少含有 10 个氨基酸残基。接头也不宜过长，以免对抗原结合部位造成干扰。目前文献中报道的接头大多含有 14～15 个氨基酸残基。有人曾尝试用 5 个氨基酸残基组成的接头，效果不佳，而长达 25 个氨基酸残基的接头仍可使 Fv 具备亲本抗体的活性。接头的氨基酸组成应使接头具备亲水性，宜于折叠，不宜有过多的侧链，以减少抗原性。用的最广泛的接头是重复出现的 4 个甘氨酸和 1 个丝氨酸（GGGGS），其中甘氨酸是分子量最小，侧链最短的氨基酸，可增加此链的柔性，丝氨酸是亲水性最强的氨基酸，可增加接头的亲水性。

ScFv 的构建在已知亲本抗体的可变区 DNA 序列时可采用化学合成的方法，也可将 V_H，V_L和接头在基因水平上拼接，方法是将接头设计在表达载体上，两端各有限制性内切酶位点供 V_H和 V_L的插入。另一种拼接法是通过 PCR 直接合成 ScFv 基因，称为重叠延伸拼接法（Splicing by overlap extension，SOE），将接头的编码序列分别设计在扩增 V_H和 V_L的引物中，并使这两个引物有 15 个碱基以上的互补序列，互为引物合成出完整的 ScFv 基因，再以两端的引物进行 PCR，即可得到足够的 ScFv 基因产物。这个方法简单易行，而且在接头两端不必因设计内切酶位点而引入不必要的氨基酸残基。V_H和 V_L基因模板的来源可以是已分离的 V_H，V_L cDNA 克隆或从杂交瘤细胞总 RNA 逆转录生成 cDNA。

(二) ScFv 的表达

各种表达系统都可以用于表达 ScFv，如：原核，酵母，植物，昆虫，哺乳类细胞等，但最常用的是大肠杆菌。从大肠杆菌表达 ScFv 可有两种方式：一种是表达为包涵体或非包涵体性不可溶蛋白，这种表达方式的产量较高，可达到细菌总蛋白的 5%～30%，但需进行变性一复性等后续工作，使其完成正确的立体结构，恢复抗体活性。另一种是分泌型表达，与 Fab 段表达的原理相同，将细菌的信号序列与 ScFv 的氨基端连接起来，使 ScFv 分子分泌到周质腔内，在周质腔内完成二硫键的形成和肽链折叠，成为有活性的单链抗体分子。这种方式可直接表达出有抗原结合活性的抗体分子，但其产量比较低，一般实验室培养条件下每升细菌的产量可达数毫克。

(三) ScFv 的应用

ScFv 以其分子小、穿透力强、廓清快、异源性低、易于大量生产等优点展示了乐观的应用前景，目前虽还未见有正式产品应用于临床，但有些已进入临床前期试用。较有前景的应用领域有以下几个方面：①导向载体：可用于临床某些疾病尤其是肿瘤的显像定位诊断和导向治疗。ScFv 廓清速度快，穿透力强，在肿瘤组织中的分布指数即肿瘤摄取量与正常组织摄取量的比值较完整抗体分子高，在放射显像时同位素排除较快，对身体危害程度小，因此是较为理想的显像定位诊断载体；②ScFv 与血液循环中抗原形成的抗原抗体复合物比完整抗体分子与抗原形成的复合物更易于被清除，因此 ScFv 很适合用于做中和抗体来阻断毒素、病毒及其他因子引起的病理发展过程。

三、胞内抗体

自杂交瘤和单克隆抗体技术问世以来，人们一直尝试将该技术用于肿瘤治疗。抗体工程的进展使得人们能够对抗体分子进行改造，研制具有高亲和力和特异结合力的基因工程抗体。同时人们对细胞内蛋白传导信号也有了更深入了解，从而可将基因工程抗体导向不同的亚细胞区室。上述两项重要进展的结合，派生出了一项全新的可阻断细胞内重要靶蛋白的胞内抗体（intracellular antibody）技术。胞内抗体技术是指应用基因重组技术在非淋巴细胞内表达具有生物活性的抗体，并通过对抗体分子进行适当修饰，使之定向分布于细胞核、细胞浆或某些细胞器中，从而特异性干扰或阻断分布于该部位的某些生物大分子的活性或加工、分泌过程，引起细胞的一系列生物过程发生改变。该技术可能对基础医学研究和肿瘤的基因治疗产生深远的影响。

胞内抗体肿瘤基因治疗是通过基因导入细胞在体内产生抗体，阻断细胞内癌相关蛋白，间接抑制肿瘤的生长。现有的研究显示，胞内抗体结合的靶蛋白及其功能主要有：① 表皮生长因子受体（EGFR）：在多种肿瘤如脑瘤、肺癌和卵巢癌中过度表达；② 白细胞介素 2 受体（IL－2R）：它是一种生长因子受体，在许多种 T 细胞和 B 细胞白血病中均有过度表达；③HER－2（erbB－2）：是一种跨膜糖蛋白，在乳腺癌和卵巢癌等许多肿瘤中均有高表达；④$p21^{ras}$蛋白：突变型的 ras 基因，广泛存在于许多人类肿瘤，其蛋白产物 $p21^{ras}$是细胞浆中的一种鸟嘌呤核苷酸结合蛋白，对细胞的生长、分化起调节

作用；⑤c－Myb 蛋白：原癌基因 c－myb 的表达产物，是癌转化的重要因素。综合现有的研究结果，其作用机制主要包括：① 抑制或稳定大分子间相互作用，改变蛋白通常所在区室；② 通过对配体的修饰使其功能丧失；③ 通过底物螯合或酶活性修饰来改变酶的功能。

四、抗体融合蛋白

将抗体分子片段与其他蛋白融合，可得到具有多种生物学功能的融合蛋白，这些融合蛋白能利用抗体的特异性识别功能将某些生物活性导向特定部位。抗体融合蛋白可有多种不同的构建方式，如将 Fv 或 Fab 段与其他生物活性蛋白融合可将特定的生物学活性导向到靶部位；在融合时可根据需要保留某些恒定区使其具备一定的抗体生物学功能；将非 Ig 蛋白与抗体分子的 Fc 段融合，可改善其药物动力学特性；将 Fv 段与其他细胞膜蛋白融合可得到嵌合受体，赋予特定的细胞以抗原结合能力。

（一）免疫导向

将毒素、酶、细胞因子等生物活性物质的基因与抗体融合，可将这些生物活性物质导向到特定的靶部位，更有效地发挥其生物学功能，减低副作用，在临床治疗中有良好的应用前景，这方面的研究相当活跃，已有许多这一类抗体融合蛋白的报道。

导向治疗是这一类融合蛋白的主要应用领域，尤其是肿瘤治疗。将抗肿瘤相关抗原的抗体与毒性蛋白融合形成的重组毒素或免疫毒素可将细胞杀伤效应引导到肿瘤部位，其抗体部分多使用 ScFv，常用抗体融合毒素有绿脓杆菌外毒素、蓖麻毒素及白喉毒素等。目前应用绿脓杆菌外毒素和蓖麻毒素的融合蛋白已进入Ⅱ期临床试用。抗体与细胞因子融合也可用于肿瘤的导向治疗，常用的细胞因子有干扰素（INF）、白介素 2（IL－2）及肿瘤坏死因子（TNF）等。

免疫导向治疗并不局限于肿瘤，可能对多种疾病起到治疗作用，如将抗纤维蛋白的 ScFv 与纤维蛋白溶酶原激活物基因拼接表达的融合蛋白可促进血栓的溶解。利用抗体融合蛋白的双重结合特性，将效应细胞与靶细胞连在一起也可以达到免疫治疗的目的，如将抗 CD3 的抗体与表皮生长因子（EGF）的基因拼接所形成的融合蛋白可以将表达有 CD3 的 T 细胞与带有 EGF 受体的肿瘤细胞连接起来，介导细胞杀伤效应。

（二）嵌合受体

将抗体的 Fv 段或 ScFv 与某些细胞膜蛋白分子融合，所形成的融合蛋白可表达于某些细胞表面，称为嵌合受体（chimeric receptor），其抗体部分行使抗原结合功能，接受特异刺激信号，由膜蛋白部分将信号传导到细胞内，产生特定的生物学效应。如将抗体的可变区基因与 T 细胞抗原受体的 α 链和 β 链的恒定区融合，将融合基因导入 T 细胞，可使 T 细胞具有该抗体的特异性，对表达相应抗原的靶细胞产生细胞杀伤效应。将 ScFv 与 T 细胞抗原受体复合体中的信息传导链，如 zeta（ζ）链融合可产生同样的效应。由于链融合受体介导的杀伤效应的识别过程是由抗体分子片段完成，从而不受主要组织相容复合物（MHC）的限制，在过继性免疫治疗中有潜在应用价值。

（三）含 Fc 的抗体融合蛋白

构建抗体融合蛋白时不仅可以利用抗体分子 Fv 段特异性结合抗原的特性，也可利用 Fc 段所特有的生物学效应功能。将某些蛋白分子与抗体 Fc 段融合可产生两种效果：一是通过该蛋白分子与其配体的相互作用将 Fc 的生物学效应引导到特定目标；二是增加该蛋白分子在血液中的半衰期，如 CD4－Fc 融合蛋白比相应的 CD4 分子在血液中的半衰期长 200 倍。CD4 是 T 细胞亚类膜表达的糖蛋白，在抗原识别中有重要功能，它是 HIV 的受体，是该病毒感染 T 细胞的必经途径，在对艾滋病的免疫防治时，由于 HIV 的变异很大，对 HIV 产生的免疫反应常难奏效，因此利用 HIV 感染 T 细胞必须经过 CD4 这一特性，阻断 HIV 对 T 细胞的识别和感染是极有潜力的治疗途径。Fc 段可对 HIV 感染的细胞介导 ADCC 活性，可激活补体，可通过胎盘对胎儿提供保护，预防垂直感染。由于 CD4 属于黏附分子范畴，CD4－Fc 融合蛋白又被称为免疫黏附素（immunoadhesin）。

第五节　抗体库技术

抗体库技术即是利用基因工程技术克隆全套人抗体重链和轻链可变区基因，并在大肠杆菌直接表达和分泌，经筛选后获得特异性抗体所获得的抗体库称为全套抗体库（repertoire antibody library），又称为组合抗体库（combinatorial antibody library）。该技术的出现标志着抗体技术发展的质的飞跃——从细胞工程抗体发展到了基因工程抗体，即第三代抗体。抗体库技术问世的基础在于两项实验技术的突破，一是 PCR 技术的发展使人们可能用一组引物克隆出全套免疫球蛋白的可变区基因；二是从大肠杆菌分泌有结合功能的抗体分子片段的成功。

一、初期的抗体库技术

美国 Scripps 研究所 Huse 等在 1989 年底首次报道了组合抗体库。其要点为：用逆转录－PCR 技术从淋巴细胞克隆出抗体轻链全套基因和重链 Fd 段全套基因，将二者分别组建到表达载体 Lc2 和 Hc2 中，得到轻链基因库和重链 Fd 段基因库，再通过 DNA 重组技术将轻链基因和 Fd 段基因随机重组于一个表达载体中，形成组合抗体库。将所得到的抗体库经体外包装后感染大肠杆菌，铺盘培养，每一个感染了噬菌体颗粒的大肠杆菌细胞由于噬菌体的增殖而裂解，所释放的噬菌体再感染周围的大肠杆菌细胞，在培养皿细菌生长层内产生噬菌斑，同时表达的 Fab 段也释放于噬菌斑内，将噬菌斑转印到硝酸纤维素膜上，可以用标记有放射性核素或过氧化物酶的抗原筛选到产生特异性抗体的克隆，得到其 Fab 段的基因。这一方法较细胞融合杂交瘤技术制备单抗有着明显的优越性：①省去了细胞融合的步骤，避免了因杂交瘤不稳定而需要反复亚克隆的繁琐程序；②扩大了筛选容量，用杂交瘤技术一般筛选能力在上千个克隆以内，而抗体库可筛选 10^6 以上个克隆；③抗体库技术直接克隆到抗体的基因，既免去了杂交瘤分泌抗体不稳定而丢失之虑，又便于进一步构建各种基因工程抗体；④抗体库技术得到的抗体可

以在原核系统表达，降低了制备成本；⑤构建抗体库时，轻链和重链可变区基因在体外组合，可产生体内不存在的轻重链配对，有可能得到新的特异性抗体。

二、噬菌体抗体库技术

上述利用λ噬菌体载体，通过噬菌斑印迹筛选抗体的抗体库技术在出现不到一年即被更为优越的噬菌体抗体库技术所取代，这是将噬菌体表面表达技术（phage display technology）应用于抗体库技术所取得的进展。

（一）表面表达

表面表达最初描述于1985年，即将蛋白分子或肽段的基因克隆到丝状噬菌体（filamentous phage，亦称单链噬菌体）的基因组DNA中，并使噬菌体表面表达该异源性分子，其主要特点是将特定分子的基因型和表型统一在同一病毒颗粒内，即在噬菌体表面表达特定蛋白，在这一核心DNA中含有该蛋白的结构基因。由于丝状噬菌体易于在大肠杆菌系统分离扩增，这个技术将选择能力和扩增能力联系到了一起，即通过与配体的结合从数量众多的多样化群体中选择出表达有相应配体的噬菌体颗粒，再通过感染大肠杆菌使选择出的噬菌体颗粒得到扩增。这一特点使表面表达成为极为有效的筛选体系。

表面表达的优越性在于具有极强的筛选功能，将序列互不相同的一组多样化的蛋白或肽链表达到噬菌体表面可得到一个库，如通过随机合成的短肽库，利用其表型与基因型的统一以及其易于扩增的特性，可很容易地从库中筛选出具有某种特异结合能力的克隆。此筛选过程如下：将库与固相化的靶分子一起温育，表达有靶分子配体的噬菌体颗粒就会结合于固相，洗去游离的噬菌体颗粒，将结合于固相的噬菌体洗脱下来，感染宿主菌进行扩增，可使特异性噬菌体得到100～1000倍的富集。这一“吸附一洗脱一扩增”的富集过程优越性是：①简便易行，一轮筛选过程1～2天即可完成；②筛选能力强，以目前技术所能产生的库容，经过多轮筛选，可选择其所含有的稀有克隆；③可直接得到特异克隆编码DNA，便于进行进一步基因改造

（二）噬菌体抗体

将抗体分子片段与噬菌体外壳蛋白Ⅲ或蛋白Ⅷ融合，使之表达到噬菌体颗粒表面，就形成了噬菌体抗体（phage antibody）。将全套可变区基因组装到表达载体内，表达子菌体表面，得到噬菌体抗体库。通过“吸附一洗脱一扩增”的富集过程，可极为有效地从噬菌体抗体库中筛选出特异性抗体的可变区基因。

（三）噬菌体抗体库技术的应用

噬菌体抗体库与初期的以λ噬菌体为载体的抗体库相比，除保持了原有的所有的优越性以外，突出的特点是极大地加强了抗体库原筛选能力。实际上其筛选能力只受到抗体库构建能力的限制。由于这些优越性，人们预测噬菌体抗体库技术将取代传统的杂交瘤单克隆抗体技术，成为制备单抗的主要手段。

噬菌体抗体库技术可用于人源化抗体的制备，由于人杂交瘤体系融合率低，稳定性差，分泌量少等低效性和人体不能随意免疫，使得采用常规杂交瘤技术制备人单抗非常困难。随着噬菌体抗体库技术的发展，不需要进行细胞融合建立杂交瘤，即可从抗体库中筛选出高亲和力的人源化抗体。这一技术的普及和推广将大大促进人源化抗体的制备。目前经噬菌体抗体库技术克隆的人抗体已陆续出现，其中抗病毒的抗体已显示出良好的应用前景。

噬菌体抗体库技术为抗体工程带来的最重要的进展之一是不经免疫制备抗体（bypassing immunization）。众所周知，欲得到高品质的抗体，首先要对动物进行有效的免疫，抗原免疫原性的强弱，免疫程序的效果，是能否得到好的单抗的第一道关口。抗体库技术还可用于改善抗体的性能，获得更具实用价值的抗体。

（卞继峰）

第七章　蛋白质和多肽药物工程

第一节　概　述

一、蛋白质多肽药物的表达系统

基因表达是一个复杂的过程，包括转录和翻译两步。不同生物细胞基因表达过程有所不同。如原核细胞转录生成的 mRNA 不经任何加工即可作为膜板，翻译合成蛋白质，而在真核细胞则需进行剪接加工才能形成 mRNA。再如，在真核细胞，有些翻译合成的多肽链需经加工修饰才能形成活性蛋白质，而原核细胞缺乏这些功能。因此，要表达真核生物蛋白质，需根据蛋白质的特点选择合适的宿主细胞。

（一）大肠杆菌表达系统

大肠杆菌表达系统是最早采用的表达系统。1978 年人类胰岛素基因在大肠杆菌获得表达，随后生长激素基因、人干扰素基因、白介素 2 基因以及人生长激素释放抑制激素相继在大肠杆菌中表达成功。

大肠杆菌表达系统的优点是：①培养方法简单，增殖快，生长成本低，具有巨大的生产潜力；②表达水平一般高于真核系统，而且其下游工艺简单，容易控制。但是大肠杆菌也存在一些不可忽略的缺点：①缺乏真核细胞特有的翻译后加工修饰功能，如糖基化、脂酰化、磷酸化等；②表达的多肽分子不能正确地折叠，以不溶性包涵体形式存在，提取纯化步骤复杂，而且从包涵体中获得的蛋白质常常没有活性；③细菌本身产生的热源、内毒素不易除去，产品纯化问题较多。上述缺点，可以通过改善细菌表达系统本身解决，许多有创意的方法已被设计出来，必要时可以在体外对表达蛋白进行加工修饰，使其具备天然蛋白质的生物学活性。

（二）酵母表达系统

酵母是单细胞真菌，是一种重要的真核表达系统。酵母表达载体中含有酵母菌的复制子、选择标记、在酵母中表达所需要的表达元件。酵母中最常用的选择标记是基于一些氨基酸或核苷酸合成酶基因突变的营养缺陷型标记，如亮氨酸（Leu）缺陷型标记和

尿嘧啶（Ura）缺陷型标记。绝大多数酵母表达载体的启动子是酵母菌自身的启动子，已发现的有磷酸甘油酸激酶（PGK）启动子、甘油醛－3－磷酸脱氢酶（GAP）启动子和醇脱氢酶（ADH）启动子等。生物学家认为，酿酒用酵母（面包酵母）是一个较好的表达系统，并已在分子水平上进行了广泛的研究。迄今为止，只有啤酒酵母（S. cervisiae）生产出的药物已用于人体，如胰岛素和乙型肝炎疫苗等。

酵母是一种低等真核细胞，他既能像其他真核细胞一样，对蛋白质进行翻译后加工修饰，进行诱导性分泌，又能像原核细胞一样可以快速增殖，容易操作，而且具有毒性小，遗传背景清楚，可根据基因表达调控机制进行高水平表达。因此酵母细胞是一种有着良好应用前景的真核基因表达系统。

（三）昆虫细胞表达系统

昆虫细胞表达系统是一类以昆虫细胞为宿主细胞的表达系统。最常用的昆虫细胞载体是昆虫杆状病毒载体，它是利用昆虫核型多角体病毒（muelear polyhedrosis Virus，Npv）的多角蛋白基因的强启动子建立起来的。昆虫细胞表达系统的优点是：①能利用信号肽使外源蛋白分泌到胞外，而且能有效地进行翻译后加工修饰，如糖基化、脂酰化和磷酸化等；②允许插入较大的外源基因，采用双启动子则可以同时插入两种外源基因；③杆状病毒具有严格的宿主专一性，对脊椎动物和植物均无致病性，也不能在非宿主细胞中繁殖和整合；④外源基因表达水平高，而且比哺乳动物细胞表达速度快。

昆虫表达系统具有易操作，安全等优点，已有 HIV 等数十种外源基因在该系统中表达成功。目前昆虫表达系统主要用于研究。所生产的有限的重组蛋白质也只用作药物设计，尚不能用于药物产业化生产。

（四）哺乳动物细胞表达系统

哺乳动物细胞表达系统是表达哺乳动物蛋白质的最佳选择。它不仅能识别和剪切外源基因中的内含子并加工成成熟的 mRNA，而且能对表达蛋白质进行翻译后加工，产生天然状态的蛋白质。

作为基因表达的宿主细胞，必须具有能在培养基中无限繁殖的特性。目前常用的哺乳动物宿主细胞，有 CHO 细胞（中国仓鼠卵巢细胞）、BHK 细胞（幼小仓鼠肾细胞）、COS 细胞（非洲金丝猴肾细胞）、Hela 细胞（人上皮癌细胞）、小鼠 NIH3T3 细胞等。哺乳动物细胞载体多为哺乳动物细胞病毒，常用的有 SV40（猴病毒 40）、EB 病毒、逆转录病毒、腺病毒、腺相关病毒、痘苗病毒等。SV40 能按正常复制周期进行复制，既可表达基因组 DNA，也可表达 cDNA，并能有条件地使外源基因获得瞬时表达或稳定（永久）表达。逆转录病毒载体能稳定地将外源基因随机整合到宿主细胞基因组内，并随细胞分裂面传代，可稳定地表达基因产物，但是这种整合只能在有 DNA 复制的细胞中进行。腺病毒及腺相关病毒既能感染增殖细胞，也能感染静止细胞。但腺病毒载体不具整合作用，因此不能持续表达外源基因产物，而腺相关病毒载体能定向整合人染色体，因面能持续表达外源基因。痘苗病毒是迄今所知结构最复杂的一类 DNA 病毒，其基因组是线状 DNA，可同时插入多个外源基因，可在多细胞中生长繁殖，不具有对真

核细胞基因转录后剪接加工能力，但表达产物可以糖基化。

（五）在转基因动物中表达

采用哺乳动物细胞表达重组蛋白质的一个突破就是，通过体外显微注射的方法将外源 DNA 注射到受精卵中产生转基因动物，外源 DNA 会随机整合到受精卵的染色质中，然后将这个受精卵植回有生殖能力的雌性动物子宫中，经过一段时间怀孕期后将产生带有外源基因的转基因动物，这种外源基因将存在于该动物的所有体细胞的染色体中。

采用转基因动物生产重组蛋白质时，其基因的表达不会受到限制，因为产生的蛋白质可储存在器官或特定区室中，不会影响动物的生理功能。最具有吸引力的转基因动物生产体系是哺乳动物的乳房，可将目的蛋白质积蓄在动物乳汁中。现已成功研制出转基因羊，利用它可以定量表达重组蛋白质，如组织纤溶酶原激活酶、α1－抗胰蛋白酶和凝血因子 IX。用转基因动物生产的蛋白质药物不久就可以上市。

二、蛋白质多肽药物研究现状

目前，利用基因工程技术生产的重组细胞因子作为生物应答调节剂（BRM）治疗肿瘤、造血障碍、感染等病症已收到良好疗效，成为新一代的药物。重组细胞因子作为药物具有很多优越之处，例如细胞因子为人体自身成分，可调节机体的生理过程和提高免疫功能，在很低剂量即可发挥作用，因而疗效显著，副作用小，已成为某些病症不可缺少的治疗手段。目前已批准生产的细胞因子药物包括 α，β，γ 干扰素，Epo，GM－CSF，G－CSF，IL－2，IL－11，EGF，bFGF 等；还有一大批细胞因子在进行临床试验，其主要适应证包括肿瘤、感染（如胆囊炎、AIDS）、造血功能障碍、创伤、炎症等（表 7－1）。

表 7－1　已批准上市的细胞因子基因工程药物

名　称	适　应　证
IFNα	白血病、Kaposi 肉瘤、肝炎、癌症、AIDS
IFNγ	慢性肉芽肿、生殖器疣、过敏性皮炎、感染性疾病、类风湿关节炎
G－CSF	自身骨髓移植、化疗导致的粒细胞减少症、AIDS、白血病、再生障碍性贫血
Gm－CSF	自身骨髓移植、化疗导致的血细胞减少症、AIDS、再生障碍性贫血
Epo	肾性贫血、癌症或癌症化疗导致的贫血、失血后贫血
IL－2	癌症、免疫缺陷、疫苗佐剂
IFNβ	多发性硬化症
IL－11	放化疗所致血小板减少症
SCF	与 G－CSF 联合应用于外周血干细胞移植
EGF	外用药治疗烧伤、溃疡
bFGF	外用药治疗烧伤、外周神经炎

另一方面，在炎症、自身免疫病、变态反应、体克等疾病时，某些细胞因子的表达

量可成百上千倍的增加，例如类风湿关节炎的滑膜液中可发现 IL－1，IL－6，IL－8 水平明显高于正常人，而这些细胞因子均可促进炎症过程，使病情加重，应用细胞因子的抑制剂有可能治疗类炎症性细胞因子水平升高的疾病（Deberts 1994）。目前，细胞因子抑制剂的研究与开发也取得了长足的进步，已有一批细胞因子抑制剂进入临床或被批准上市，包括人源化的抗细胞因子或其受体的抗体、可溶性细胞因子受体、细胞因子受体拮抗剂、细胞因子和毒素的融合蛋白等，例如抗 TNF 受体的抗体、可溶性 TNF 受体、IL－2－白喉毒素融合蛋白等已在近年被批准上市，前两者治疗类风湿关节炎，后者治疗 T 细胞淋巴瘤，均收到非常明显的疗效。

第二节 激素类药物

一、胰岛素

（一）结构与理化特性

胰岛素（Insulin）是一种蛋白质类激素，由胰岛素的 β 细胞合成。人类胰岛素由 A，B 两条多肽链构成，A 链含有 21 个氨基酸残基，B 链含有 30 个氨基酸残基，这两条链由二硫键连接，在 A 链中还含有一个链内二硫键。用碱或还原剂使二硫键断裂，胰岛素失活。蛋白水解酶能使胰岛素水解失活，因此，胰岛素不能口服。

胰岛素在胰岛 β 细胞初合成时是一条由 102 个氨基酸残基组成的单链多肽，称为前胰岛素原（preproinsulin）。前胰岛素原穿过内质网膜进入腔内，随即切去前面由 16 个氨基酸残基组成的信号肽，生成 86 肽的胰岛素原（preproinsulin），并输送到高尔基体储存。当胰岛素原分泌时，在蛋白水解酶作用下，将连接 A，B 链的连接肽（C 肽）切下，释放出 A，B 两条短肽，通过两个二硫键连成成熟的胰岛素。

胰岛素的分子量约为 6000，在不同的 pH、离子强度和蛋白浓度下，可以单体、二聚体、六聚体等状态存在。胰岛素的自聚合不仅与它在体内储存、运输和生物活性相关，而且与它在临床上的应用密切相关。事实上，胰岛素以含 Zn^{2+} 的六聚体在胰脏中存在，由胰脏分泌进入血液后，迅速解离为二聚体和单体，其在体内的活性单位是单体。作为药物制剂在临床应用的注射胰岛素是六聚体，进入体内后也经历与上述相似的过程。

（二）生物学活性

胰岛素是一种重要的蛋白质激素，不仅对机体的糖代谢具有关键作用，而且具有促生长功能。胰岛素的主要生物学作用是促进组织、细胞对葡萄糖的摄取和利用，加速糖原合成、抑制糖异生，促进葡萄糖转变为脂肪，从而使血糖降低。胰岛素缺乏时，血糖浓度升高，如超过肾糖阈，尿中将出现糖，引起糖尿病。故胰岛素是治疗糖尿病常用药物。胰岛素还能抑制脂肪酶的活性，减少脂肪的分解，并能促进蛋白质的合成。

（三）重组人胰岛素

人胰岛素是采用基因工程技术生产的第一个经批准而用于人体的药物。以往治疗糖尿病都用猪或牛胰岛素，虽然它们在人体有生物学活性，但其氨基酸序列和人胰岛素不完全相同，用后有的病人会产生抗体，影响继续用药的效果，偶然还会出现严重的免疫反应。而重组人胰岛素与天然产品相同，不存在免疫源性问题。用基因工程生产人胰岛素不宜从染色体上取得基因着手，因为这样做的最终产物是前胰岛素原，还含有不编码介入序列，所以采用了分别合成编码胰岛素 A，B 链的核苷酸序列，A 链需要 63 个核苷酸，B 链需要 90 个核苷酸，两条链的起始端都加上甲硫氨酸密码子 ATG，将这两条链分别插入表达载体与 β－半乳糖苷酶基因中，分别转化大肠杆菌后，表达得到胰岛素链与 β－半乳糖苷酶羟基端融合的蛋白质，提取融合蛋白，并用溴化氰处理。因 β－半乳糖苷酶蛋白中含有其他甲硫氨酸，故被切成许多小肽而被删除，而胰岛素 A，B 链中均无甲硫氨酸，不被切割，经过提纯获得天然胰岛素多肽，将两者混合，即可形成具有活性的重组胰岛素。这一探索性研究现已进行改进，即以反转录的 cDNA 插入产生单一 β－半乳糖苷－胰岛素融合蛋白方式，只要一步切割即可释放出成熟胰岛素，这个改进方法已被用来生产人重组胰岛素。200L 培养液的胰岛素产量达 10g，相当于约 450kg 胰脏的提取量。

（四）人胰岛素的定点改造

如前所述，目前作为药物制剂在临床用的注射胰岛素是六聚体，进入体内后要解聚为单体才能有效发挥生物学作用。六聚体胰岛素解聚为单体的速度，可影响其降血糖的速度。如果通过分子改造使注射胰岛素进入体内后，解聚过程延长，即可制备长效胰岛素；相反，使注射胰岛素主要以单体形式存在或使其从六聚体解聚为单体的过程缩短，即可制备速效胰岛素。蛋白质工程技术的发展，使这一设想有可能得以实现。至今利用定点诱变技术对胰岛素分子进行改造已获得 100 多种定点改造的胰岛素突变体，现举例作一概要介绍。

1. 速效胰岛素　重组突变体 B9（Ser→Asp）人胰岛素（以下简称 B9Asp）是通过蛋白质工程途径制备的一种速效胰岛素。它将处于二聚体界面上的中性、短链残基 B9Ser 突变为酸性、较长链的 Asp，使其在中性 pH 条件下主要以单体形式存在，注入体内后无需解聚进入血液，从而产生速效效应。实验证明，在中性条件下，B9Asp 浓度达 1mmol/L 时仍为单体，而且保证了 26％的 FFC（自由脂肪细胞实验）活力和 79％的降血糖（MBGA）活力。

2. 长效胰岛素　Novo 研究所利用定点诱变技术对人胰岛素进行定点改造，获得［A21Gly，B27Arg，B30Thr 酰胺］突变体，是目前已知在血液中半衰期最长（35.5hr）的一种长效胰岛素，称 Novosol Basal，已进入临床试验。它的长期效应主要在于 B 链羧端 Thr 被酰胺化和 B27 的 Thr 突变为带正电荷的 Arg，使其等电点上升（从 pH5.4 升至 pH6.8），而 A21 位的突变使其在酸性溶液中也不会脱酰胺，保证了它的稳定性，这样制成 pH3.0 的可溶制剂，注射后能在生理 pH 条件下形成结晶，延缓

了吸收，半衰期因而延长，有利于低剂量胰岛素基础水平的维持。

3. 高效胰岛素 高效胰岛素类似物的研究开展较早。1987 年由 Schwartz 等利用定点诱变技术将人胰岛素 B10 的 His 突变为 Asp，获得 B10 Asp 突变体。该突变体的体外生物活力和受体结合能力均显著升高，约为天然人胰岛素的 5 倍。[B10Asp（去 B26～B30）B25Tyr 酰胺] 人胰岛素是目前为止人们获得的活力最高的胰岛素类似物，它的体外活力是人胰岛素的 11.7 倍。

总之，随着胰岛素蛋白质工程的发展，人们已经获得大量的胰岛素类似物，对这些类似物的研究，加深了人们对胰岛素结构与功能的关系以及胰岛素作用机理的了解。同时，人们也获得了一批速效和长效胰岛素，它们的治疗学性质比天然胰岛素优越，为改善糖尿病的治疗提供了光明前景。另一方面，由于这些类似物对人体来说，属于突变蛋白，可能引起免疫反应，某些类似物还具有较强的促生长作用。因此，在广泛应用于临床之前，也需要对它们的安全性和生理反应作长时间的动物和人体实验。

二、生长激素

（一）结构与生物学特性

人生长激素（human growth hormone，hGH）由脑下垂体分泌，系单链多肽，由 191 个氨基酸残基组成，分子量约为 21.5KD。人生长激素对不同组织具有多种效应，一般能增强全面生长。在儿童时期缺乏生长激素会导致侏儒症，如能定期注射生长激素，可促进其生长达到或接近正常高度。虽然人、牛和猪生长激素的氨基酸排列极为相似，但只有人类和其他灵长类的生长激素对人有活性。因此，对侏儒症患儿的治疗，不能以牛或猪的生长激素代替。

（二）重组人生长激素

基因工程的发展，为重组 hGH 提供了必要理论和技术基础。重组 hGH 的生产将为患儿提供安全、有效、足量且便宜的药物。最初生产的重组 hGH 是采用构建一个杂合基因来完成的。通过反转录获得编码 1－191 氨基酸的 cDNA，用 EcoR I 切除 1～23 氨基酸编码序列，再与人工合成的编码 1－23 氨基酸的 DNA 相连接，这个编码序列被重组人邻近细菌启动子的质粒载体。hGH 与胰岛素一样，以含有信号肽序列的较大的前体蛋白方式产生。因为人的信号序列不被细菌的分泌机制所识别，所以这个 hGH cDNA 的 5′端要用合成序列来代替，使宿主菌能生产一种近乎正常结构的成熟 hGH，只是其氨基端有一个额外的起始甲硫氨酸，其活性与天然 hGH 相同。产生这种蛋白质的另一方式是构建能分泌蛋白质的生产菌株，方法是将蛋白质的编码序列与细菌分泌蛋白质的信号序列相连接，以前激素（pre－hormone）方式来生产，在细菌周质腔（periplasmic space）内积聚，周质腔内的蛋白酶将信号肽切除，用低渗裂解外膜而获得不含起始甲硫氨酸且与天然 hGH 完全相同的蛋白质。由于周质腔内的杂蛋白含量较少，纯化较简易。用细菌生产的人生长激素治疗侏儒症，获得了较好的效果。450L 工程菌培养物的产量可以相当于 6 万个人脑下垂体的提取量。

（三）人生长激素的定点改造

hGH 既能与生长激素受体结合又能与催乳素（prolactin）受体结合。为了避免治疗过程中产生副作用，有必要通过分子改造，使 hGH 只能与生长激素受体结合，而不与催乳素受体结合。因为生长激素分子与生长激素受体结合的区域与能与催乳素受体结合区域部分重叠，因而有可能通过分子改造减少其与催乳素激素受体结合的可能。

hGH 三维结构与功能研究发现，His－18，His－21 和 Glu－174 是 Zn^{2+} 结合部位，当 Zn^{2+} 与这些氨基酸侧链结合后，hGH 与催乳素受体具有高亲和性。利用定点诱变技术，改变上述部位的氨基酸序列即可降低 hGH 与催乳素受体的结合。目前已经获得了仅能与 hGH 受体结合而不与催乳素受体结合的 hGH 突变株。尽管此项研究极为有趣，但是这种人生长激素类似物能否用于临床，尚需进行大量的研究。

三、促红细胞生成素

人促红细胞生成素（Epo）是由 190 个氨基酸残基组成的高度糖基化蛋白质，其中 27 个氨基酸为信号肽，成熟蛋白质由 163 个氨基酸残基组成，其糖基对于 Epo 的生物活性至关重要，因而基因工程的 Epo 不能利用大肠杆菌表达系统，只能利用哺乳动物表达系统如 CHO 细胞生产。Epo 含有 2 对二硫键，等电点为 4.5，具有一定的酸稳定性和热稳定性。

Epo 主要由肾脏产生，约 10%～15%的 Epo 由肝脏产生，其功能是刺激造血祖细胞分化为红细胞，维持外周血的正常红细胞水平。在肾功能衰竭时，出现 Epo 的产生障碍，导致贫血。动物体内实验证明人重组 Epo 能够明显刺激红细胞的产生，防止肾功能衰竭导致的贫血。

CHO 细胞生产的人重组 Epo 已被批准上市，作为治疗肾衰导致贫血的首选药物，在全世界销售了数十亿美元，成为最成功的基因工程药物之一。人重组 Epo 亦可治疗一些非肾性原因导致的贫血，如慢性感染、炎症、放化疗等。此外，Epo 还可明显减少化疗后的输血量和减轻贫血，这表明 Epo 的适应证可进一步扩大。

第三节　细胞因子类药物

细胞因子类药物是由细胞产生和分泌的，能影响细胞生长、分化和增殖的一类蛋白质，它们不仅能参与胚胎发育、组织器官的发生及组织自我更新的调控作用，而且与炎症和创伤愈合以及恶性转化等现象有关。目前已知的细胞因子种类很多，预计还将有更多的细胞因子被发现。由子细胞因子具有临床治疗的实际意义，其开发前景诱人。目前世界上有数十家公司在进行这方面的研究，已有 α－干扰素（IFN－α）、粒－巨噬细胞集落刺激因子（GM－CSF）、巨噬细胞集落刺激因子（M－CSF）等获美国食品及药物管理局（FDA）批准。

一、干扰素

干扰素（interferan，IFN）是由多功能细胞因子组成的异源家族，具有广泛的生物学活性，对细胞生长、分化以及免疫功能等均有重要的调节作用。IFN 是由英国科学家 Isaacs 和 Lindemann 于 1957 年发现，由于产量极少，其生物学功能及临床应用研究进展较慢，自从工程菌发酵生产干扰素成功后，对干扰素的研究才有了较大进展。

（一）结构与分类

根据结构和来源干扰素分为三类，即白细胞干扰素（IFN－α）、成纤维细胞干扰素（IFN－β）和免疫干扰素（IFN－γ），其中 IFN－α 又分为多种亚型。根据一级结构，IFN－α 分为 2 个亚族（subfamily），分别称为 IFN－αⅠ和 IFN－αⅡ，目前常用的 IFN－α 为 IFN－αⅠb，IFN－αⅡa 和 IFN－αⅡb。IFN－αⅠb 由 166 个氨基酸残基组成，相对分子量为 19329，含有 5 个 Cys，形成两个二硫键，其余一个 Cys 的存在使其易形成二聚体，降低活性，其等点电介于 5～6 之间。IFN－αⅡa 由 165 个氨基酸残基组成，分子量为 19219，含有 4 个 Cys，形成两个二硫键，等点电介于 5～6 之间。IFN－αⅡb 与 IFN－αⅡa 只相差一个氨基酸，即 23 位的 Cys 改为 Arg，性质也相仿。

IFN－β 由 166 个氨基酸残基组成，与 IFN－α 之间约有 25%～30%的同源性，含有 3 个 Cys，其中 Cys41 和 Cys141 形成二硫键，在 80 位氨基酸残基有一个 N－糖基位点。IFN－γ 的一级结构与 IFN－α 和 IFN－β 无同源性，它由 143 个氨基酸组成，另有 23 个氨基酸的信号肽。成熟 IFN－γ 无 Cys，因而无二硫键的形成，在 28 和 100 位氨基酸残基各有一个 N－糖基化位点。

（二）生物学活性

干扰素是免疫系统抵御外源侵袭的一类细胞因子，其主要作用有：①通过影响细胞内代谢过程，合成多种抗病毒蛋白，从而抑制病毒等细胞内微生物的增殖；②通过干扰细胞分裂周期抑制细胞增殖；③通过作用于 CTL、单核 β 巨细胞、NK 细胞、T 淋巴细胞和 B 淋巴细胞而进行免疫调节；④增加组织相容性抗原的表达。临床研究表明，干扰素对某些疾病有较好的疗效，如 IFN－α 能有效地治疗乙肝、丙肝、单纯疱疹病毒感染、慢性病毒感染性疾病等；IFN－β 能有效地治疗病毒引起的带状疱疹，对乳腺癌、肾癌、恶性黑色素瘤等也有一定的作用；IFN－γ 对类风湿性关节炎有一定疗效。

（三）重组人干扰素

干扰素具有高度的种属特异性，所以起初临床使用的干扰素都用人细胞制备，产量极低，限制了对干扰素的研究和应用。1973 年 Cohen 首先采用体外重组技术，将两种不同 DNA 重组，并用大肠杆菌表达，获得了人重组干扰素，使大量廉价制备单组分干扰素成为可能。1980 年，利用基因工程获得了 α－干扰素和 β－干扰素的 cDNA，1982 年又获得了 γ－干扰素的 cDNA。将这些 cDNA 分别与高表达质粒载体重组成重组子，转导人大肠杆菌中进行表达，获得了高产量的干扰素。如大肠杆菌表达 α－干扰素，每

升菌液中含有 2.5 亿 U 的 α－干扰素，相当于从 100L 人血中获得的提取量，利用杆状病毒载体将 α－干扰素在家蚕虫表达，每 mL 体液中也可获得 2 亿 U 的产物。基因工程干扰素的产生推动了干扰素的产业化生产，1987 年，三种干扰素的生产开始工业化，并大量进入市场。中国预防医学科学院病毒研究所、卫生部上海生物制品研究所、卫生部长春生物制品研究所和中国药品生物制品鉴定所等单位联合研制了基因工程 IFN－αIb，并由长春生物制品研究所投入工业化生产，成为我国第一个进入工业化的基因工程药物。

二、白细胞介素

白细胞介素（Interleukin，IL）是一组具有介导白细胞相互作用的细胞因子，在免疫系统中发挥重要生理功能。自 1979 年第一个白细胞介素被命名后，发现和克隆新的白细胞介素，一直是国际免疫学研究的热点，至 1996 年已发现了 18 种，分别命名为 IL_{1-18}，它们的来源及生物学活性见表 7－2。

表 7－2　几种白细胞介素的生物活性

名称	曾用名称	产生细胞	生物学活性
IL－1	淋巴细胞活化因子	巨噬细胞、成纤维细胞	激发全身抗感染防御反应，激活各种免疫活性细胞
IL－2	T 细胞生长因子（TCGF）	活化 T 细胞	促进 T 和 B 细胞的增殖和分化，诱导和促进多种细胞毒性，抑制胶质细胞的生长
IL－3	多功能集落刺激因子（multi－CSF）	活化 T 细胞	促进造血干细胞和 T 细胞增殖，促进肥大细胞和粒细胞的增殖和分化
IL－4	B 细胞分化因子（BCGF－1）	活化 T 细胞	促进 B 和 T 细胞增殖，调节 T 和巨噬细胞功能
IL－5	T 细胞代替因子（TRF）	活化 T 细胞	促进 B 细胞的生长与分化，促进嗜酸粒细胞增殖与分化
IL－6	B 细胞刺激因子－2（BSF－2）	淋巴细胞、单核细胞、成纤维细胞	诱导 B 细胞增殖分化，产生 Ig 并支持多功能干细胞的增殖，提高 NK 细胞活性，刺激造血细胞
IL－7	淋巴细胞生成素（LPO）	骨髓及胸腺基质细胞	促进 B 和 T 细胞前体的增殖，促进胸腺细胞的生长
IL－8	粒细胞活化因子（NAF）	单核/巨噬细胞、血管内皮细胞	激活中性粒细胞，促进血管生成
IL－9	T 细胞生长因子Ⅲ（TCGFⅢ）	活化 T 细胞	促进 T_H 细胞增殖

名称	曾用名称	产生细胞	生物学活性
IL－10	细胞因子合成抑制因子（CSIF）	B细胞	刺激肥大细胞及其祖细胞，促进B细胞增殖
IL－11	－	骨髓基质细胞	刺激巨噬细胞成熟，增加血小板数量
IL－12	细胞毒淋巴细胞成熟因子（CLMF）	B细胞	诱导多种细胞因子如IFN－γ等，激活NK细胞和T细胞
IL－13	－	活化T细胞	抑制细胞因子的分泌和表达，刺激B细胞增殖和CD23的表达
IL－14	高分子B细胞生长因子（HMN－BCGF）	T细胞	诱导激活的B细胞增殖，抑制丝裂原刺激的B细胞分泌Ig
IL－15	－	成纤维细胞、角质细胞、内皮细胞和单核/巨噬细胞	与IL－2的作用相似，能刺激T细胞的增殖和分化、激活NK细胞以及促进造血细胞的增殖和分化
IL－16	淋巴细胞趋化因子	成纤维细胞、淋巴细胞	对$CD4^+$ T细胞、单核/巨噬细胞和嗜酸粒细胞有化学趋化和免疫调节作用，诱导IL－2R表达
IL－17	细胞毒淋巴细胞相关抗原8（CTLA－8）	T细胞	促进造血祖细胞增殖，刺激释放多种炎症细胞因子
IL－18	IFN－γ诱导因子（IGIF）	单核/巨噬细胞	增强NK细胞活性，诱导T_H细胞产生IFN－γ

IL－2是第一个被发现的白细胞介素，1983年已从cDNA文库中克隆成功，目前它已在大肠杆菌、酵母及某些哺乳动物细胞系中得到表达。IL－2是一个由133个氨基酸残基组成的多肽，其N端有一个20个氨基酸残基组成的信号肽，当其分泌时被切除。IL－2有3个Cys，形成一个二硫键，其中125位的Cys处于游离状态，它可能与IL－2的稳定性有关。有人利用定位诱变技术将Cys^{125}的密码TGT转换成TCT或GCA，使Cys^{125}转换成丝氨酸或丙氨酸，IL－2的生物活性不受影响，而稳定性有所增加。基因工程IL－2已被美国FDA批准上市，用于治疗肾细胞瘤。

1999年11月至2000年底，仅仅一年时间又发现至少5个新的白细胞介素（$IL_{19\sim23}$）。这些新白细胞介素的发现均采用了计算机克隆的技术，即利用商业化的EST数据库，通过计算机同源性分析，克隆、表达及功能分析，最终获得新的白细胞介素。这种技术路线实质上是采用了反向生物学的原理，从基因到蛋白质再到体外功能和体内功能研究的途径来发现新的生物活性分子。

白细胞介素的种类较多，结构各异，但是这些细胞因子均由白细胞分泌，是调节哺乳动物细胞生长与分化的细胞网络的重要组成部分。它们通过复杂的细胞因子网络，介

导免疫应答与炎性反应，调节造血功能，并对多种细胞的生长与分化起重要的调控作用。大量研究表明，白细胞介素对肿瘤、感染性疾病及免疫缺陷等疾病均具有较好的疗效。其中 IL－2 已被美国 FDA 批准上市，用于治疗肾细胞瘤，其他的白细胞介素如 IL－3，IL－4，IL－6，IL－10，IL－11 和 IL－12 在国外已进入临床试验。目前我国卫生部已批准试生产基因工程白细胞介素－2，正在进行开发性研究的有 IL－2，IL－3，IL－4 和 IL－6 等。

三、集落刺激因子

集落刺激因子（colony－stimulating factors，CSF）是一类促进造血干细胞增殖，分化和成熟的一类蛋白质因子。集落刺激因子的种类很多，根据其来源分为粒细胞集落刺激因子（G－CSF）、粒/巨细胞集落刺激因子（GM－CSF）和巨噬细胞集落刺激因子（M－CFS）等。

（一）结构与理化性质

1. 人 G－CSF　人 G－CSF 基因位于第 17 号染色体的 q^{21-22} 区，大约由 2500 个核苷酸组成，含有 4 个内含子和 5 个外显子。G－CSF 基因转录后产生的前体 mRNA，由于内含子 2 的剪切拼接位置的差异，产生 2 种 mRNA，编码相差了 3 个氨基酸残基的 2 种蛋白。

cDNA 克隆表达获得两种 G－CSF，分别由 204 和 207 个氨基酸残基组成，在穿越细胞膜分泌 G－CSF 到细胞外时，其 N－端由 30 个氨基酸残基组成的信号肽段被酶切掉，分别得到含 174 个氨基酸残基和 177 个氨基酸残基的成熟蛋白，后者在前者的 35 和 36 位氨基酸残基之间插入 Glu，Ser 和 Val 3 个氨基酸残基，活性大大下降。G－CSF 共有 5 个 Cys，在 Cys^{36} 与 Cys^{42}，Cys^{74} 与 Cys^{64} 之间形成两个二硫键，从而使其折叠形成高级结构，是 G－CSF 保持生物活性所必须的。G－CSF 相对较稳定，4℃以下活性可维持 1 年以上，在 pH2～10 时均较稳定。

2. 人 GM－CSF　人 GM－CSF 基因是位于第 5 号染色体，全长约为 2.5kb，含 4 个外显子和 3 个内含子，其表达产物是由 127 个氨基酸残基组成的蛋白质分子，有 2 个 N－糖基化部位和 3 个 O－糖基化部位，含 4 个半胱氨酸，Cys^{54} 与 Cys^{98}，Cys^{88} 与 Cys^{121} 间形成 2 个二硫键。由于糖基化程度不同，天然的人 GM－CSF 的相对分子量一般为 14500～32000，糖基化程度高的 GM－CSF 其比活性反而低。GM－CSF 在 4℃下，稳定性达 6 个月，pH＜2 时可保持稳定性，对胰蛋白酶、胰凝乳蛋白酶均不敏感。

（二）生物学活性

G－CSF 能促进粒系族细胞向成熟中性粒细胞增殖和分化，并维持其存活功能。G－CSF 的应用前景广阔，对于各种原因引起的粒细胞减少症均有显著疗效。

GM－CSF 能刺激粒细胞和巨噬细胞的生长和分化，诱导早期红细胞、巨核细胞、嗜酸性粒细胞前体细胞的生长和增殖，刺激内皮细胞的趋化性及增殖。也能激活成熟血细胞的功能和黏附性，抑制中性细胞从感染部位移开，促进内皮细胞的有丝分裂，所以

GM－CSF 在机体防御中具有重要作用，在临床上有很大的应用价值。

（三）重组 G－CSF 和 GM－CSF

人和鼠的 G－CSF 以及 GM－CSF 均具有较高同源性，已经从小鼠纯化获得天然 G－CSF 和 GM－CSF，但因含量太少，纯度不高，利用基因工程生产 G－CSF 和 GM－CSF 已势在必行。Gough 等于 1984 年根据从肺组织中纯化的 GM－CSF 序列合成了寡核苷酸作为基因探针，从鼠肺中克隆出了鼠 GM－CSF 的 cDNA；1985 年又从 ConA 活化的 T 细胞系 LB_3 的 cDNA 文库中克隆出了鼠 GM－CSF 的 cDNA；人 GM－CSF 的 cDNA 是从人白血病细胞系 Mo 或人 T 淋巴细胞系 T_7 的 cDNA 文库中筛选得到的。随后 Dune 等人完成了基因组 DNA 克隆工作。至今，GM－CSF 基因已分别在大肠杆菌、酵母菌及哺乳动物细胞中表达成功。由大肠杆菌表达的 GM－CSF 不带糖链，由哺乳动物细胞表达的 GM－CSF 其 N－糖基化位点及 O－糖基化位点均被糖基化。GM－CSF 是第一个被克隆的造血因子，也是第一个利用重组 DNA 技术生产的造血因子。现在，rhGM－CSF 已广泛应用于临床，是一种治疗造血系统功能异常的有效药物。

G－CSF 的基因克隆也获得了成功，并分别在大肠杆菌和真核细胞得到表达，大肠杆菌表达的 rhGM－CSF 虽不能糖基化，但其活性不变。应用基因工程方法生产 G－CSF 产率为 20～50mg/L 培养液，而且重组 G－CSF 很容易被纯化，用凝胶过滤和疏水柱层析纯化效率为 20%～50%。

（刘传华　卞继峰）

第八章　蛋白质分离纯化技术

生物样品中，除含有目的蛋白质外，还含有其他蛋白质、多糖、脂类和核酸等混杂物质，要研究某一种蛋白质的结构、理化性质及其功能或制备某种蛋白制剂，必须进行分离纯化。根据蛋白质的分子结构和理化性质的差异，蛋白质分离纯化方法不同。破碎组织细胞，提取蛋白质混合物，利用盐析法、有机溶剂沉淀法等可对蛋白质进行粗提；电泳和色谱技术则可对粗提蛋白进行进一步的分离纯化，而且可对纯化蛋白质进行分析鉴定和制备。本章从原理上对蛋白质分离纯化中常用的电泳技术和色谱技术作以概要介绍。

第一节　蛋白质分析分离中的电泳技术

一、概　述

蛋白质是一种两性电解质，在一定 pH 条件下可解离成带电离子，在电场作用下可向与其本身所带电荷相反的电极移动，这种现象称为电泳（electrophoresis）。由于不同蛋白质的等电点不同，在同一 pH 缓冲液中所带电荷性质及电荷量不同，加上各种蛋白质分子的黏度和分子量不同，在同一电场作用下移动的方向和速度不同，因此，利用电泳技术可对蛋白质进行分离、纯化。电泳技术是由 Tiselius 于 1937 年首先建立，1948 年由于电泳方法的改进而被广泛接受和重视。随着科学技术的迅猛发展，电泳技术也不断被完善和发展，特别是电脑和网络技术的发展，使电泳的控制及电泳结果的分析更为精确和深入，因而其应用也就更广泛，已成为蛋白质研究领域不可缺少的技术。

（一）电泳的分类

电泳的种类很多，根据电泳的方法、支持物、方向等的不同，电泳的分类也不一样。根据有无支持物分为自由界面电泳和区带电泳。自由界面电泳是利用胶体溶液的溶质颗粒经过电泳后，在溶液和溶剂之间形成界面，从而达到分离的目的。区带电泳也称为支持电泳，是样品在惰性支持物上进行电泳的过程，因为有支持物的存在减少了界面之间的扩散程度和干扰的发生，而且多数支持物还具有分子筛的作用，更增加了电泳的分辨率，加上区带电泳简单易行，成为目前应用的主要的电泳技术。

根据电泳使用的技术可分为显微电泳，免疫电泳，密度梯度电泳，等电聚焦电泳等；根据电泳的方向又分为水平电泳和垂直电泳；根据所用支持物的不同又分为纸电泳、醋酸纤维素膜电泳、琼脂糖凝胶电泳及聚丙烯酰胺凝胶电泳等；根据电泳系统的连续性还可分为连续性电泳和不连续性电泳等。

（二）电泳的基本原理

电泳是带电粒子在电场中移动的现象。当把一个带电荷（q）的颗粒放入电场时，便有一个力（F）作用于其上。F 的大小取决于颗粒静电荷及其所处的电场强度（E），它们的关系是：

$$F=E \cdot q \qquad ①$$

由于 F 的作用，使带电颗粒在电场中向一定方向泳动。此颗粒在泳动过程中还受到一个相反方向的摩擦力（f）阻挡，当这两种力相等时，颗粒则以相等速度（v）向前移动，即：

$$v=F/f= E \cdot q/f \qquad ②$$

式中 f 为摩擦系数。根据 stoice 公式，阻力大小取决于带电颗粒的大小、形状及所用介质的黏度，即：

$$f=6\pi r\eta \qquad ③$$

式中 r 为颗粒半径、η 为介质黏度。该公式系指球形颗粒所受的阻力。把③代入②则：

$$v= E \cdot q / (6\pi r\eta) \qquad ④$$

从公式④看出，带电颗粒在电场中泳动的速度与电场强度和带电颗粒的净电荷量成正比，与颗粒半径和介质黏度成反比。蛋白质是两性电解质，在一定 pH 下，可解离成带电荷的离子，在电场作用下可以向与其电荷相反的电极泳动，泳动速度主要取决于蛋白质分子所带电荷的性质、数量及颗粒的大小和形状。由于各种蛋白质的等电点（pI）不同，分子量不同，在同一 pH 缓冲溶液中所带电荷不同，故在电场中的泳动速度也不同。利用此性质，可将混合液中不同的蛋白质分离开，也可用其对样品的纯度进行鉴定。

（三）影响电泳速度的因素

1. 电场强度　电场强度是指每厘米的电位降，也称电势梯度（电位梯度）。电场强度对电泳速度起着决定作用。电场强度愈高，带电颗粒泳动速度愈快。根据电场强度大

小，又将电泳分为常压电泳和高压电泳，用高压电泳分离样品需要的时间比常压电泳短。

2. 缓冲溶液的 pH 值　缓冲溶液的 pH 值决定带电颗粒解离的程度，亦即决定其净电荷的量。对于蛋白质而言，缓冲溶液的 pH 值离等电点越远，所带净电荷的量越多，泳动速度越快，反之则越慢。因此，当要分离某一蛋白质混合物时，应选择一个合适的 pH 值，使各种蛋白质所带电荷的量差异大，有利于彼此分开。

3. 缓冲溶液的离子强度　溶液的离子强度也能影响电泳的速度。溶液的离子强度增高，缓冲液负载的电流增强，样品所负载的电流则降低，使带电颗粒的泳动速度降低；若离子强度过低，则缓冲能力差，往往会因溶液 pH 值的变化，影响颗粒的泳动速度。

4. 电渗　在电场中液体对于固体支持物的相对移动称为电渗。电渗是由于支持介质中某些基团解离并吸附溶液的正离子或负离子，使靠近支持物的溶液相对带电而造成的。例如滤纸的纤维间具有大量孔隙，其中的一些基团可能解离成正离于，能吸附溶液中的负离子，而与纸相接触的水溶液带正电荷，液体便向负极移动。在电场作用下，液体向负极移动时，可携带颗粒同时移动。例如在 pH8.6，血清蛋白进行纸电泳时，γ一球蛋白与其他蛋白质一样带负电荷，应该向正极移动，然而它却向负极方向移动，这就是电渗作用的结果。所以电泳时颗粒泳动的表观速度是颗粒本身的泳动速度与由于电渗影响而携带的移动速度两者的加和。若电泳方向与电渗方向相同，则其表观速度将比泳动速度快，若二者方向相反，则其表观速度将比泳动速度慢。

5. 焦耳热　在电场中根据热量 $Q=A^2\Omega$，电泳时会产生热量，使电泳系统的温度升高，并且电阻随温度升高而下降，因此如果电压保持不变，那么温度升高会引起电流的升高，并且促使支持介质上溶剂的蒸发，为了消除这些不利影响使电泳的可重复性增高，可使用经过稳定的电源装置，进行稳压或稳流电泳；还可用一个密闭的电泳槽以减少缓冲液的蒸发；在电泳槽中加设一个冷却系统，起到外加冷却的作用。

二、聚丙烯酰胺凝胶电泳

聚丙烯酰胺凝胶电泳（polyacrylamide gel electrophoresis，PAGE）是在 1959 年由 Ornstei 和 Darvis 所创造，至 1964 年他们从理论上进一步阐述，在技术上也作了改进，该方法操作简单，分辨率高。如用醋酸纤维素薄膜电泳只能将血清蛋白分离成为 5～8 个组分，而聚丙烯酰胺凝胶电泳则能使之分离成为 30 个以上的组分，因此这个方法被迅速推广，成为生物学研究的常规技术。

根据凝胶的形状，PAGE 可以分为圆盘状电泳（Disc page electrophoresis）和垂直或水平板状电泳（Slad page electrophoresis）。垂直电泳或水平电泳的同一块胶板上可以同时分离数个样品，重复性较圆盘电泳为好，同时平板型电泳结合等电聚焦、免疫电泳等进行两相分析，还利于进行显影等显示已分离的各个组分，故近年来板状电泳的应用更为广泛。

（一）基本原理

PAGE是一种以聚丙烯酰胺凝胶为支持介质的区带电泳。聚丙烯酰胺凝胶由单体（monomer）丙烯酰胺（acrylamide Acr）和交联剂（crosslinker）N，N－甲叉双丙烯酰胺（N，N－methylene－bis acrylamide，Bis）聚合交联而成。聚丙烯酰胺凝胶具有三维网状结构，电泳颗粒通过网状结构的空隙时受到摩擦力的作用，摩擦力的大小与样品颗粒的大小呈正相关。可见，样品蛋白在电场作用下受到两种作用力，即静电引力和摩擦产生的阻力，因而大大提高了电泳的分辨能力。

1. 凝胶的聚合　Arc和Bis在有自由基存在时，即可聚合成凝胶。引发产生自由基的方法有化学法和光学法

（1）化学聚合：化学聚合的催化剂一般采用过硫酸铵（ammonium persulfate，AP）加速剂是脂肪族的叔胺，如四甲基乙二胺（tetramethyl ethylenediamine，TEMED）、三乙醇胺和二甲基氯丙腈等，其中以TEMED为最好。当过硫酸铵被加入Acr，Bis和TEMED的水溶液时，立即产生过硫酸自由基，该自由基再激活TEMED，随后TEMED作为一个电子载体提供一个未配对电子，将丙烯酰胺单体活化。活化的丙烯酰胺彼此连接，聚合成含有酰胺基侧链的脂肪族多聚链。相邻的两条多聚链之间，随机地通过Bis交联起来，形成三维网状结构的凝胶物质。

（2）光聚合：通常以核黄素为催化剂，不加TEMED即能聚合，但加入可加速聚合。光聚合通常需要痕量氧的存在，核黄素经光解形成无色基，再被氧化成自由基，从而引发聚合作用。一般光聚合用于制备大孔胶，化学聚合用来制备小孔胶。

2. 凝胶用量计算　聚丙烯酰胺凝胶孔径的大小取决于反应体系中凝胶总浓度和交联度。总浓度通常用$T\%$表示，即100mL凝胶溶液中含的Arc和Bis的总克数；交联度通常用$C\%$表示，即交联剂（Bis）占单体和交联剂总量的百分数。凝胶浓度和交联度的计算公式如下：

$$T\%=\frac{(a+b)}{m}\times 100\% \qquad C\%=\frac{b}{(a+b)}\times 100\%$$

式中a代表单体Arc的重量（g），b代表交联剂Bis的重量（g），m代表溶液的体积（mL）。

通常凝胶的孔径、透明度和弹性随着凝胶浓度的增加而降低。分离不同分子量蛋白质的混合物时，只有选择适宜浓度的凝胶才能奏效。常用于分离血清蛋白的标准凝胶是指浓度为7.5%的凝胶。用此胶分离大多数生物体内的蛋白质，电泳结果一般都满意。当分析一个未知样品时，常常先用7.5%的标准凝胶或用4%～10%的梯度凝胶试验，以便选择到理想浓度的凝胶。当分离物的分子量已知时，可参考表8－1选择适宜的凝胶浓度。

表 8－1　选择聚丙烯酰胺凝胶浓度参考值

样品	分子量范围	适宜凝胶浓度（T%）
蛋	$<10^4$	20～30
白	$1-4\times10^4$	15～20
质	$4\times10^4-1\times10^5$	10～15
	$10^5-5\times10^5$	5～10
	$>5\times10^5$	2～5
核	$<10^4$	15～20
酸	10^4-10^5	5～10
	$10^5-2\times10^5$	2～2.6

3. 分离效应　聚丙烯酰胺凝胶电泳分为连续性电泳和不连续性电泳两个系统。连续性电泳系统的凝胶浓度一致，凝胶中的 pH 值及离子强度与电泳槽液的相同。不连续电泳系统存在四个不连续性：①凝胶层的不连续性，通常含有 3 种性质不同的凝胶（见表 8－2）；②缓冲液离子成分的不连续性；③pH 值的不连续性；④电位梯度的不连续性。由于上述四种不连续性，电泳时会产生 3 种物理效应，即样品的浓缩效应、凝胶的分子筛效应以及电荷效应，因此，聚丙烯酰胺凝胶电泳具有很高的分辨能力。

表 8－2　不连续电泳三层凝胶的性质

	Tris－HCl	凝胶浓度	凝胶孔径
样品胶	pH6.7	3%	大（大孔凝胶）
浓缩胶	pH6.7	3%	大（大孔凝胶）
分离胶	pH8.9	7.5%	小（小孔凝胶）

（1）浓缩效应：由于不连续电泳系统的凝胶层不连续，缓冲液离子成分的不连续性以及由此造成的电位梯度的不连续性，使得样品在电泳开始时得以浓缩。在浓缩胶中其缓冲液的 pH 为 6.7，在该 pH 下，HCl 几乎全部解离；甘氨酸的等电点为 pH6，故甘氨酸的解离度很小；血清蛋白质的等电点多数为 pH5 左右，其解离度在 HCl 和甘氨酸之间，在这个系统中含有 Cl^-、Pr^-（蛋白质离子）和 $NH_2-CH_2-COO^-$（甘氨酸离子 g）三种带负电荷的离子，在电场中这三种离子在大孔胶内的有效迁移率的大小顺序依次为：

$m_{Cl}\cdot\alpha_{Cl}>m_{Pr}\cdot\alpha_{Pr}>m_g\cdot\alpha_g$　　（m：迁移率，α：解离度，$m\cdot\alpha$：有效迁移率）

Cl^-称为快离子，或前导离子（Leading ion），甘氨酸称为慢离子或尾随离子（Trailing ion）。当电泳刚开始时，由于三种胶中都含有 Tris－HCl 缓冲液，都含有快

离子 Cl^-，电泳槽中只含有慢离子。电泳开始后，因快离子有最大的有效迁移率，很快超过蛋白质和甘氨酸离子，而使其后面形成一个低电导即高电势梯度的区域。这种高电势梯度使蛋白质和慢离子在快离子后面加速移动，致使高、低电势梯度区之间形成一个快速移动的界面，由于蛋白质的有效迁移率介于快、慢离子之间，蛋白质离子后有甘氨酸离子向前推，前有 Cl^- 阻挡，蛋白质就聚集在 Cl^- 和甘氨酸离子之间，被浓缩成薄薄的一层。

另一方面，蛋白质分子在大孔胶中受到的阻力小，移动速度快，进入小孔胶时遇到的阻力大，速度减慢。由于凝胶层的不连续性，在大孔与小孔胶的界面处就会使样品浓缩，区带变窄。

(2) 电荷效应：在一定的 pH 环境中，各种离子所带电荷不同，其迁移率也不同，不同蛋白质分子的等电点不同，其所带的表面电荷也各不相同，因此它们的迁移率不同。经电泳后，各种蛋白质根据其迁移率的大小依次排列成一条条的区带。浓缩胶和分离胶中均存在这种电荷效应。但是在经过 SDS 处理后的蛋白质，由于 SDS 的电荷掩盖了蛋白质本身所带的电荷，因而在 SDS－PAGE 中蛋白质的分离不依赖电荷的差别。

(3) 分子筛效应：蛋白质在电场作用下泳动，受到两种作用力，即静电的引力和介质的阻力。在聚丙烯酰胺凝胶电泳中静电的引力与其他电泳一样，主要取决于蛋白质颗粒的自身带电性状，但所受到的介质的阻力则决定于凝胶孔径的大小，凝胶具有三维结构，凝胶的浓度不同其网孔的孔径大小不同，根据分离的对象可以调节凝胶浓度，使网孔孔径与分离对象的分子大小处于相匹配的状态，蛋白质分子通过凝胶时，受到的阻力与分子大小成正相关，所以分子大小不同的蛋白质即使所带电荷相同，自由迁移率相等，但在电泳一段时间后，也能彼此分开，这就是分子筛效应。

(二) 聚丙烯酰胺凝胶电泳的优点

以聚丙烯酰胺凝胶为支持介质进行蛋白电泳，可根据被分离物质分子大小及电荷多少来分离蛋白质，具有以下优点：①聚丙烯酰胺凝胶是由丙烯酰胺和 N，N′甲叉双丙烯酰胺聚合而成的大分子。凝胶是带有酰胺侧链的碳－碳聚合物，没有或很少带有离子的侧基，因而电渗作用比较小，不易和样品相互作用；②由于聚丙烯酰胺凝胶是一种人工合成的物质，在聚合前可调节单体的浓度比，形成不同程度网孔结构，其空隙度可在一个较广的范围内变化，可以根据要分离物质分子的大小，选择合适的凝胶成分，使之既有适宜的网孔，又有比较好的机械性质。一般说来，含丙烯酰胺 7%～7.5%的凝胶，机械性能适用于分离分子量范围在 1 万至 100 万的物质，1 万以下的蛋白质则采用含丙烯酰胺 15%～30%的凝胶，而分子量特别大的可采用含丙烯酰胺 4%的凝胶，大孔胶易碎，小孔胶则难从管中取出，因此当丙烯酰胺的浓度增加时可以减少双丙烯酰胺，以改进凝胶的机械性能；③在一定浓度范围内聚丙烯酰胺对热稳定，凝胶无色透明，易观察，可用检测仪直接测定；④丙烯酰胺是比较纯的化合物，可以精制，减少污染。

(三) 影响凝胶聚合的因素

聚丙烯酰胺凝胶是电泳支持介质，其质量直接影响分离效果，对影响凝胶聚合的因

素应特别注意。

1. 形成凝胶的试剂的纯度　丙烯酰胺是形成凝胶溶液中的最主要成分，其纯度的好坏直接影响凝胶的质量。此外，丙烯酰胺和 bis 中可能混杂有能影响凝胶形成的杂质，如丙烯酸、线性高聚丙烯酰胺、金属离子等，这些物质也能影响凝胶的聚合质量，影响电泳的结果，应予以充分注意。对丙烯酰胺，最好是选择质量好，达到电泳纯级的产品。过硫酸铵容易吸潮，而潮解后的过硫酸铵会渐渐失去催化活性，故过硫酸铵溶液需新鲜配制。

2. 凝胶浓度　每 100mL 凝胶溶液中含有单体和胶联剂的总克数称凝胶浓度，常用 T%表示。可供选择的凝胶浓度范围为 3%～30%，凝胶浓度增加后聚合速度将加快，因此可适当减少催化剂的用量。凝胶浓度的大小，还会影响凝胶的质量和网孔的大小。凝胶网孔的大小与总浓度相关，总浓度越大，孔径相对变小，故在用聚丙烯酰胺凝胶电泳分离蛋白质时，应根据需要选择凝胶浓度，在一般情况下，大多数生物体内的蛋白质采用 7.5%浓度的凝胶，所得电泳效果往往是满意的，故称由此浓度组成的凝胶为“标准凝胶”。凝胶浓度过大，凝胶透明度差，而且因其硬度及脆度较大，容易破碎；而凝胶浓度太低时，形成的凝胶稀软，不易操作。

3. 温度和氧气的影响　聚丙烯酰胺凝胶聚合的过程也受温度的影响，温度高聚合快，温度低聚合慢，一般以 23℃～25℃为宜。大气中的氧能淬灭自由基，使聚合反应终止，所以在聚合过程中要使反应液与空气隔绝，最好能在加激活剂前对凝胶溶液脱气。

三、SDS 聚丙烯酰胺凝胶电泳

蛋白质在聚丙烯酰胺凝胶中电泳时，它的迁移率取决于它所带净电荷以及分子的大小和形状等因素。如果加入一种试剂消除电荷、形状等因素的影响，使电泳迁移率只取决于分子的大小，就可以用电泳技术测定蛋白质的分子量。1967 年，Shapiro 等发现阴离子去污剂十二烷基硫酸钠（SDS）具有这种作用并建立了 SDS－聚丙烯酰胺凝胶电泳（SDS－PAGE）。其原理是当向蛋白质溶液中加入足够量 SDS 和巯基乙醇，SDS 能破坏蛋白质分子间共价键，使蛋白质变性而改变原有的构象，特别是强还原剂巯基乙醇的存在，使蛋白质分子内的二硫键还原，从而保证蛋白质与 SDS 结合形成带负电荷的蛋白质－SDS 复合物。蛋白质－SDS 复合物所带的 SDS 负电荷的量大大超过了蛋白质分子原有的电荷量，因而掩盖了不同种蛋白质间原有的电荷差别，另外，蛋白质－SDS 复合物的形状是类似于“雪茄烟”形的长椭圆棒，不同蛋白质的 SDS 复合物的短轴长度是恒定的，约为 18 埃，而长轴的长度则与蛋白质分子量大小成比例，这样的蛋白质—SDS 复合物，在凝胶中的迁移率，不再受蛋白质原有的电荷和形状的影响，而取决于分子量的大小。因此，SDS－PAGE 可以按蛋白质的分子大小的不同将其分开。大多数肽链迁移率与其分子量呈对数线性关系，若用已知分子量的一组蛋白质作图绘制标准曲线，在同样条件下检测未知样品，就可从标准曲线推算出未知样品的分子量（见图 8－1）。

SDS－PAGE 主要用于蛋白质分子量的测定、蛋白质混合组分的分离和蛋白质亚基

组分的分析等方面。当蛋白质经 SDS—PAGE 分离后，如果把各种蛋白质组分从凝胶上洗脱下来并且把 SDS 除去，还可用于氨基酸测序，酶解图谱以及抗原特性等方面的研究。SDS—PAGE 后的凝胶还可直接用于转移印迹分析。

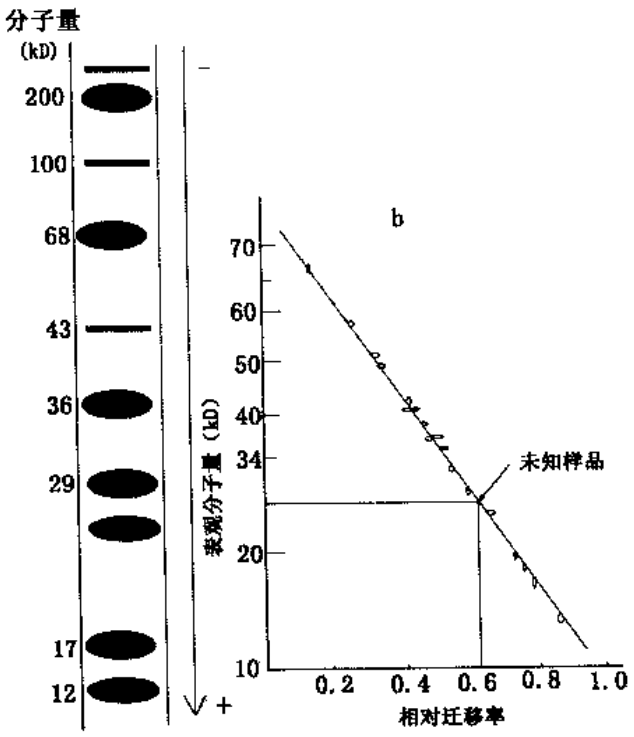


图 8—1 SDS—PAGE 测定蛋白质分子量

四、等电聚焦电泳

等电聚焦（isoelectric focusing，IEF）是 1966 年由瑞典科学家 Rible 和 Vesterberg 建立的一种高分辨率的蛋白质分离分析技术。它利用蛋白质分子或其他两性分子等电点的不同，在一个稳定、连续的线性梯度中进行蛋白质的分离。

（一）基本原理

蛋白质是由不同种类的 L—α-氨基酸按不同的比例以肽键相连而构成的，构成蛋白质的一些氨基酸在一定 pH 的溶液中是可解离的，从而可带有一定的电荷。构成蛋白质的所有氨基酸残基上所带正负电荷的总和便是蛋白质的净电荷。在低 pH 时，蛋白质的净电荷是正的，在离 pH 时，其净电荷是负的，若在某一 pH 的溶液中，某蛋白质的净电荷为零，则此 pH 值即为该蛋白质的等电点（pI）。由于不同蛋白质有着不同的氨基酸组成，所以蛋白质的等电点值取决于其氨基酸的组成，是一个物理化学常数。每一种蛋白质都有其特定的氨基酸组成，所以各种蛋白质的等电点不同，因此可以利用电泳技术，根据蛋白质等电点的差异对其进行分离分析。

等电聚焦就是在电泳槽中放入载体两性电解质，当通以直流电时，两性电解质即形成一个由阳极到阴极逐步增加的 pH 梯度。当蛋白质放进此体系时，靠近阳极侧的蛋白质处于酸性环境中，带正电荷向负极移动；靠近阴极的蛋白质处于碱性环境中，带负电荷向正极移动，最终都聚焦于与其等电点相当的 pH 位置上，形成不同的蛋白质区带。等电聚焦的优点是：①有很高的分辨率，可将等电点相差 0.01～0.02pH 单位的蛋白质分开；②一般电泳由于受扩散作用的影响，随着时间和泳动距离加长，区带越走越宽，而电聚焦能抵消扩散作用，使区带越走越窄；③由于这种电聚焦作用，不管样品加在什么部位，都可聚焦到其等电点位置，很稀的样品也可进行分离；④可直接测出蛋白质的等电点，其精确度可达 0.01pH 单位。电聚焦技术的缺点是：一是要求用无盐溶液，而在无盐溶液中蛋白质可能发生沉淀，二是样品中的成分必需停留于其等电点，不适用在等电点不溶或发生变性的蛋白质。

（二）pH 梯度的建立

1. 载体两性电解质　产生 pH 梯度的方法有两种：①用两种不同 pH 的缓冲液互相扩散，在混合区形成 pH 梯度，这是人工 pH 梯度；②利用载体两性电解质（carrier ampholytes）在电场作用下自然形成 pH 梯度，称为天然 pH 梯度，该方法是常用的方法。

天然 pH 梯度的原理是 Svensson 于 20 世纪 60 年代初期提出的。其形成过程是，当电解硫酸钠的稀溶液时，在阳极聚焦硫酸而在阴极聚焦氢氧化钠。若将一些两性电解质放入电泳槽中，则它们在阳极的酸性介质中就会得到质子而带正电，在阴极的碱性介质中则失掉质子而带负电，这样就会受其附近的电极所排斥而向相反方向移动。设其中有一个酸性最强的两性电解质甲，当它由阴极逐渐接近阳极的硫酸时，就会失去电荷而停止运动，甲所在的位置就是它的等电点。另有一个等电点稍高于甲的物质乙，当向阳极运动靠近甲时，它不能超过甲，因为那里低于它的等电点。于是乙将带正电荷而向阴极移动，它只能排在甲的阴极侧。假如有很多两性电解质，它们就会按照等电点由低到高的顺序依次排列，形成一个由阳极向阴极逐步升高的平稳的 pH 梯度，此梯度的进程取决于两性电解质的 pH 值、浓度和缓冲性质。防止对流的情况下，只要电流稳定，这个 pH 梯度将保持不变。

为保证等电聚焦分离分析蛋白质的效果，载体两性电解质应具备如下条件：①在等电点处必须有足够的缓冲能力，以便能控制 pH 梯度，而不致被样品蛋白质或其他两性物质的缓冲能力改变 pH 梯度的进程；②在等电点必须有足够高的电导，以便使一定的电流通过，而且要求具备不同 pH 值的载体有相同的电导系数，使整个体系中的电导均匀；如果有局部电导过小，就会产生极大的电位降，从而其他部分电压就会太小，以致不能保持梯度，也不能使应聚焦的成分进行电迁移，达到聚焦；③分子量要小，便于与被分离的高分子物质用透析或凝胶过滤法分开；④化学组成应不同于被分离物质，不干扰测定；⑤应不与分离物质反应或使之变性。总起来说，当一个两性电解质的等电点介于两个很近的 pH 值之间时，它在等电点的解离度大，缓冲能力强，而且电导系数高。

2. 等电聚焦的支持介质　目前常用的支持介质有聚丙烯酰胺凝胶、琼脂糖凝胶和葡聚糖凝胶。其中聚丙烯酰胺凝胶是等电聚焦电泳分析中最广泛采用的支持介质。凝胶在等电聚焦电泳中除了起支持介质作用外，尚能防止已聚焦分子的扩散、对流，从而使蛋白质样品在凝胶上可分离出更多致密的区带。

等电聚焦电泳的方式有多种，大致可分为垂直管式、毛细管式、水平板式及超薄水平板式。这些方式各具其特点。目前在样品分析中趋向子选用超薄水平板式，它具有分析样品多，两性电解质用量少、结果重复性好等优点。

（三）等电聚焦中应注意的事项

1. pH 梯度的选择，可先在宽 pH 范围内载体两性电解质中进行分析 IEF。当了解到目的蛋白质的 pI 值后，再用窄 pH 范围的载体两性电解质进行分析或制备。

2. IEF 如在宽 pH 范围载体两性电解质内进行，为了克服在中性区域形成纯水区

带，可适当添加中性载体两性电解质。

3. 为防止电泳过程中 pH 梯度的衰变，一般电流降低达最小而恒定时尽快结束 IEF。

4. pH 梯度（pI）测定：①Rotofer 制备电泳可分管测定收集液的 pH 值；②凝胶 IEF 后，可分段切割凝胶，用 3～5 倍体积的蒸馏水或 10mmol KCl 浸泡该凝胶，从凝胶浸出液中测定 pH 值，或用微电极直接测定凝胶表面 pH 值；③薄层等电聚焦后，可用微电极检测凝胶表面 pH 值根据蛋白质分布的 pH 值即能确定该蛋白质的 pI。

5. IEF 聚焦过程中由于蛋白质泳动到某一区段，而该区段刚好是该蛋白质的 pI，造成蛋白质的沉淀出现絮凝现象，为了解决此问题，可在样品中添加脲、TritonX－100，NP—40 或其他一些非离子型表面活性剂。

五、双向凝胶电泳

双向凝胶电泳是一种由任意两个单向凝胶电泳组合而成的，即在第一向电泳后再在与第一向垂直的方向上进行第二向电泳。其基本原理一般与组成它的两个单向电泳的基本原理相同。早在 1975 年 O'Farrell 首先建立了等电聚焦/SDS－聚丙烯酰胺双向凝胶电泳（IEF/SDS－PAGE），至今仍不失为双向凝胶电泳首选的组合方式。在这项组合中，IEF 为第一向电泳，是基于蛋白质的等电点不同用等电聚焦法分离，第二向则按分子量不同用 SDS－PAGE 分离，把复杂的蛋白混合物中的蛋白质在二维平面上分开。近年来经过多方面改进，已成为研究蛋白质组的最有价值的核心方法。

虽然双向凝胶电泳的基本原理与单向电泳基本相同，但在操作上却有所差别。第一向 IEF 通常为盘状电泳，等电聚焦系统内含有高浓度的尿素和非离子型去污剂 NP－40，而且溶解蛋白质样品的溶液除含有尿素和 NP－40 外，还含有二硫苏糖醇（dithiothreitol）。这些试剂本身并不带电荷，不影响各蛋白质组分原有的电荷量和等电点，但能破坏蛋白质分子内的二硫键，使蛋白质充分变性和肽链舒展，从而有利于蛋白质分子在温和条件下与 SDS 充分结合，以提高第二向电泳的效果。为提高分辨率，20 世纪 80 年代开始采用固定化 pH 梯度胶，克服了载体两性电解质阴极漂移等许多缺点而得以建立非常稳定的可以随意精确设定的 pH 梯度。由于可以建立很窄的 pH 范围（如 0.05U/cm），对特别感兴趣的区域可在较窄的 pH 范围内做第二轮分析，从而大大提高了分辨率。此种胶条已有商品生产，因此基本上解决了双向凝胶电泳重复性的问题。这是双向凝胶电泳技术上的一个非常重要的突破。

第一向电泳后蛋白质得到了初步分离，凝胶条中所含的蛋白质区带，是第二向电泳的样品，因此，进行第二向电泳前，要将第一向电泳后的凝胶条包埋在第二向电泳的凝胶板中。因为两向电泳的分离系统不同，需将第一向电泳后的凝胶预先在第一向电泳分离系统中振荡平衡，然后再包埋于第二向电泳凝胶板的电泳起始端，然后按照 SDS－PAGE 的方法进行电泳。

第二向 SDS－PAGE 有垂直板电泳和水平超薄胶电泳两种做法，可分离 10～100kD 分子量的蛋白质。其中灵敏度较高的银染色法可检测到 4ng 蛋白，最灵敏的还是用同位素标记，20ppm 的标记蛋白就可通过其荧光或磷光的强度面测定。用图像扫描仪、莱

赛密度仪、电荷组合装置可把用上述方法得到的蛋白图谱数字化，再经过计算机处理，去除纵向和横向的曳尾以及背景底色，就可以给出所有蛋白斑点的准确位置和强度，得到布满蛋白斑点的图像，即所谓“参考胶图谱”。蛋白质组研究的主要困难是对用双向凝胶电泳分离出来的蛋白，进行定性和定量的分析。最常用的方法是先把胶上的蛋白印迹到 PVDF（polyvinylidene difluoride）膜上后再进行分析，确定它们是已知还是未知蛋白。现在的分级分析法是先做快速的氨基酸组成分析，也可先做 4～5 个循环的 N 末端微量测序，再做氨基酸组成分析；结合在电泳胶板上估计的等电点和分子量，查对数据库中已知蛋白的数据，即可作出判断。有文献报道，N 末端 4 个残基序列的数据就可以给出很多的信息而得到相当准确的结果。如再结合 C 末端序列，判断结果的准确性会更高。对分离得到的蛋白质作进一步的确切鉴定需要有足够数量的纯蛋白，电泳时蛋白质已经过了高度纯化。现在一块胶板可允许上到高达 mg 数量级的样品，因此每个分离的蛋白斑点可有 μg 数量的蛋白，这样使本来是微量的蛋白也可被鉴定。

蛋白质的翻译后修饰和加工，是指在肽链合成完成后进行的化学反应，如磷酸化、羟基化、糖基化、二硫键形成，以及最近发现的蛋白质自剪接等等，可能有一百种以上。翻译后修饰和加工对蛋白质的正常生理功能是必需的，它们的变化往往和疾病的发生有关。用双向凝胶电泳可以进行翻译后修饰的研究，如用^{32}P 标记可以研究磷酸化蛋白的变化。双向凝胶电泳中常可发现的蛋白质拖曳现象，很可能是一个蛋白的不同翻译后修饰产物所造成的。拖曳图像变化在疾病诊断上可能提供重要的信息。

双向凝胶电泳技术当前面临的挑战是：① 低拷贝蛋白的鉴定，人体的微量蛋白往往还是重要的调节蛋白。除增加双向凝胶电泳灵敏度的方法外，最有希望的还是把介质辅助的激光解吸/离子化质谱用到 PVDF 膜上，但当前的技术还不足以检出拷贝数低于 1000 的蛋白质；② 极酸或极碱蛋白的分离；③ 极大（>200kD）或极小（<10kD）蛋白的分离；④ 难溶蛋白的检测，这类蛋白中包括一些重要的膜蛋白；⑤ 得到高质量的双向凝胶电泳需要精湛的技术，因此迫切需要自动二维电泳仪的出现。

六、毛细管电泳

20 世纪 80 年代很多科学家致力于发展微量的，没有支撑介质的液相毛细管电泳，自 1989 年有商品化仪器供应后，毛细管电泳的应用研究更是日新月异，开始主要用子蛋白质和多肽方面，以后应用在核酸、糖、维生素、药品检验、无机离子、环保等各个领域，和 HPLC 相映成为分析方法中互补的技术。

（一）基本原理

1. 毛细管电泳　所谓毛细管电泳是在内径为 25～100μm 的石英毛细管中进行电泳，毛细管中填充了缓冲液或凝胶。与平板凝胶电泳相比，毛细管电泳可以减少焦耳热的产生，这主要是由于毛细管内径细，因此表面积与体积比大，易于扩散热量；另外电泳时电阻相对大，即使选用较高电压（可高至 30kV）仍可维持较小的电流。毛细管电泳通常在高电场下进行，可以缩短分析时间并且提高分辨率。

2. 毛细管电泳仪　目前生产毛细管电泳仪的厂商不下数十家，其基本结构都是由

毛细管、高压电源、电极及电极液、在线检测器、恒温装置、样品盘、数据收集和处理系统等组成（见图 8—2），毛细管是由熔融石英（fused silica）制成的，内径通常为 25～75μm，外径为 350～400μm，在其外壁涂有一层聚亚胺保护层，使得毛细管有一定柔性不易折断。在检测窗口需将这层保护材料去除（可用小火灼烧几毫米，再用乙醇将黑灰抹去）。去除了保护层的毛细管非常脆弱容易折断，操作时要格外小心。

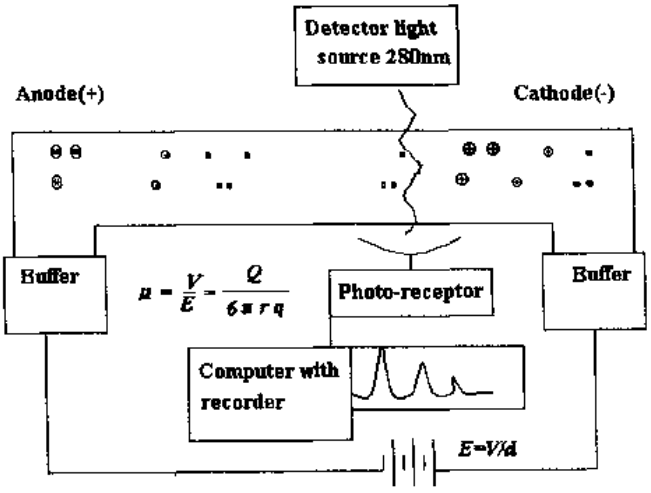


图 8—2 毛细管电泳仪结构图

毛细管电泳仪的高压装置一般可允许电压高至 30kV，最大电流为 200～300 mA，并且可改变其正、负极方向。在电泳过程中，毛细管两端及电极插至电极液，电极液通常与管中的缓冲液一致。正、负两电极连接至高电压装置。进样时样品盘移动，使毛细管的进样端及此端的电极（通常是正极）准确插入样品管中，给予一定电场或压力后，样品被吸入毛细管内。然后再将毛细管移至电极液中开始电泳。

常用的检测方式是紫外一可见光吸收。检测器位于距样品盘约为毛细管总长的 2/3～4/5 处，对毛细管壁内部分进行光聚焦。此外，荧光、激光诱导荧光、质谱等检测方法也被应用于毛细管电泳。

虽然毛细管的内径很细有助于散热，但是由于 1℃的温度变化会引起缓冲液粘度出现 2%～3%的差异，因此仍需借助有效的散热装置，如简单的风扇以及含有循环液体的冷却系统。理论上讲液体冷却的效果更好，但如果空气流动速度达到 10 m/秒时，也足以能够将电泳产生的热量散去。

3. 电渗 在水性条件下大多数固体表面都带有多余（额外）的负电荷，这可能是由于酸碱平衡造成的离子化和/或固体表面吸附了离子基团所致。构成毛细管的熔融石英由硅醇（—SiOH）组成，它在水溶液中以—SiO—形式存在，为了维持电荷平衡，溶液中的正离子吸附至石英表面形成双电子层，当在毛细管两端施加了电压后这一层正离子趋向负极移动，并带动毛细管中的溶液以液流（bulk solution）的形式移向负极（见图 8—3），这一现象称为电渗（electroosmosis）或电渗流（electroendosmotic flow，EOF）。由于毛细管的表面积与体积之比大，且电泳时使用了高电压，电渗流在毛细管电泳中常常起着不可忽视的作用。

电渗流的大小可以下面两个公式表示：

$$V_{EO} = (\varepsilon\zeta/\eta)\ E$$
$$\mu_{EOF} = (\varepsilon\zeta/\eta)$$

其中 V_{EOF} 为电渗流速率；μ_{EOF} 为电渗流泳动力；ζ 为电势能；ε 为双电子常数；η 为溶液的粘度。ζ 势能是由毛细管壁的表面电荷决定的，面表面电荷依赖于溶液的 pH，因此电渗流和电泳缓冲液的 pH 密切相关。在高 pH 环境中硅醇基团易于去质子化，电渗流大于低 pH 溶液时的电渗流。ζ 势能和缓冲液的离子强度也有关，增大溶液的离子强度即减少 ζ 势能，也就降低了电渗作用。

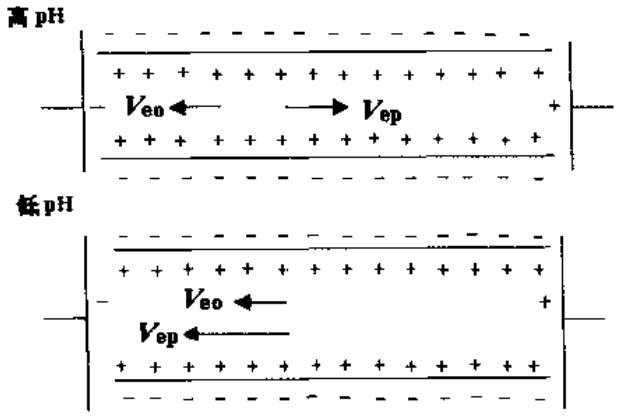


图 8—3　电渗流示意图

Veo 为电渗流方向，Vep 为样品分子泳动方向

电渗作用特点之一是液体沿着毛细管壁均匀流动，因面其前沿是平头的，不同于高效

液相色谱中泵推动的液体流动呈圆弧状。另一特点是电渗使得携带不同电荷的分子朝一个方向移动。不带电荷的中性分子也能随着电渗流一起移动。

电渗的上述两个优点在毛细管电泳分离中起着有益的作用，但是在实验中有时能够发现选择过高 pH 缓冲液进行电泳时电渗流太快以至不能将被分析物分开；而选择过低 pH 缓冲液时，毛细管内壁的负电荷也能吸附被分析物中带正电荷的分子。测定电渗流的大小可采用中性标准物质，如二甲亚砜（DMSO）、氧化甲基硅以及丙酮。

（二）毛细管电泳分离模式

与 HPLC 相似，毛细管电泳根据不同的分离机制，有多种分离模式。

1. 自由溶液毛细管电泳（free solution capillary electrophoresis，FSCE）或称毛细管区带电泳（capillary zone electrophoresis，CZE），毛细管中充满了电泳缓冲液后，由于电渗的作用，分析物中带正、负电荷及中性的分子均能向一个方向移动，但是不同的中性分子没有自身的泳动力，都和电渗流共迁移，它们之间不能分开。

由于 FCSE 具有与反相 HPLC 不同的分离机理，且其理论塔板系数不亚于后者，可成为后者的一项互补技术，尤其对于亲水多肽的分离分析具有一定的优越性。在某种程度上，FCSE 较 HPLC 更方便、有更多的选择性。只需改变电泳缓冲液的条件，如 pH、离子强度等，不必更换毛细管。在合适的电泳条件下，毛细管电泳可以分辨多肽的细微结构差异，如 C—端的酰胺化与去酰胺化，单个氨基酸残基的替换，氨基酸组成相同而序列不同的多肽也能区别开，甚至表面电荷不同而一级结构相同的多肽也有可能分离开。还有报道用 FSCE 帮助确定多肽是否存在二硫键。

2. 微团电动毛细管色谱（micellar electrokinectic capillary chromatography，MECC）MECC 在某种程度上与反相 HPLC 类似通常在电泳缓冲液中添加一定浓度的表面活性剂，这类分子一端为亲水基团，其余部分为疏水结构。当表面活性剂的浓度高于其临界微团浓度时，表面活性剂分子就会发生聚集，形成微团或胶束。其中疏水部分在内部，亲水基团在表面，被称为假固相。根据假固相和溶液相中的分配不同各种分子得到分离。在应用 MECC 模式时常可加入有机修饰剂，如：甲醇、乙腈、异丙醇等。可降低溶质和微团间的疏水吸附，也可减少形成微团的疏水作用。由于 MECC 可以同时分离带电荷与不带电荷的分子，目前被较多的应用于氨基酸衍生物、肽谱分析以及合成多肽纯度鉴定等。

3. 毛细管筛分电泳（capillary sieving electrophoresis，CSE）此种电泳模式的特点是毛细管内填充了多聚物作为分离介质，对分子的泳动起着阻碍作用。其作用机理类似于凝胶的筛网作用，大分子较小分子所受的阻力大，泳动速度减慢。这种分离模式尤其适合于荷质比差异很小而大小不同分子的分离，这类分子应用 FSCE 无法分析。CSE 在 DNA 研究中有着广泛的应用，在蛋白质研究中主要用来代替传统的平板 SDS—聚丙烯酰胺凝胶电泳。其优点在于：①缩短分析时间，CSE 只需 15 分钟即可达到与凝胶电泳同样的分离效果；②CSE 可以在线检测并进行定量；③自动化程度高，省却手工电泳操作的麻烦。但是 CSE 每次只能分析一个样品，另外，目前 CSE 只能用于分子量 1 万至 20 万范围内蛋白质的分离，尚不可进行多肽的分析。

4. 毛细管等电聚焦（capillary isoelectric focusing，CIEF）与普通的凝胶等电聚焦一样，毛细管内的 pH 梯度也是由两性电解质形成。在实验中需选用内壁中性共价涂层的毛细管，阳极端至检测器有效分离部分（离检测窗口差几厘米）充满两性电解质溶液，样品溶液夹在其中以避免直接接触阳极液。毛细管其余部分（检测器至阴极）为阴极液。毛细管两端分别插入阳极和阴极液中，其中毛细管中的溶液和溶液中均含一定浓度的可溶性甲基纤维素作为支持介质。两性电解质载体和样品在电场中聚焦至电流趋于零。此时可采用真空抽吸、压力推动或者在电极液中加入盐，以改变已经形成的 pH 梯度等方式驱使蛋白质条带移动并逐一经过检测窗口。CIEF 不仅可以取代常规的凝胶等电聚焦，还具有前述的诸多优点。CIEF 可以在线分析多肽的等电点。CIEF 具有分辨率高，可用于分析差异较小的异构体。对于较稀的蛋白溶液（低至 0.001μg/μL）也可以分析。

七、聚丙烯酰胺凝胶印渍转移电泳

印渍转移电泳又称电转移印渍（electrotransfer blotting），它是以凝胶电泳为基础检测蛋白质的一项重要生化技术。其基本原理是利用低压电泳的直流电使凝胶电泳的分离区带转移到有一定韧性并且化学惰性的分子支持物上，然后再利用各种检测方法对这些印渍进行分析鉴定，从而从众多的分离区带中检验出所需的某一组分或某一片段。印渍转移电泳是在 Southern 印渍技术的基础上发展起来的。Southern E. M 于 1975 年首先创立了 DNA 印渍技术，即将经限制性核酸内切酶降解后的各种 DNA 片段进行琼脂糖凝胶电泳分离，然后经毛细管虹吸作用将凝胶上的 DNA 区带转移到硝酸纤维素膜上，最后与标记探针杂交即可检测出所需 DNA 片段。该技术被命名为 Southern Blotting（Southern 印渍）。此后将此法应用于 RNA，称为 Northern Blotting。1979 年 Towbin 等开始研究蛋白质分离区带转移，他们设计了一种电转移印渍装置，在凝胶电泳后，将带有蛋白质分离区带的凝胶与硝酸纤维素膜紧贴在有孔转移框架内，组成“夹心饼”。然后置于低压高电流的直流电场中，以电驱动为转移方法的蛋白质印渍法称为 Western Blotting，又称为蛋白质印渍。1982 年，Reihart 等人将等电聚焦电泳后的蛋白质分离区带也以电驱动为转移方式进行了印渍。这一方法称为 Eastern Blotting，也属于蛋白质印渍范畴。由于电转移印渍与凝胶电泳都是在直流电场下进行，所使用的装置也很相似，因此，人们通常把电转移印渍称为转移电泳或印渍转移电泳。

印渍转移电泳是当代分析和鉴别生物大分子最有意义的技术之一，具有如下优点：①与毛吸管作用方式的 Southern Blotting、Northern Blotting 相比，电转移速度快，印渍效果好；②利用该技术可以从同一凝胶上获得多张相同的印渍转移复本，在同一张固相纸膜上可进行多次分析鉴定；③该技术灵敏度高、试剂用量少、操作简单、所需仪器设备也比较简单。因此，聚丙烯酰胺凝胶印渍转移电泳在生物化学、分子生物学、分子遗传学、免疫学以及临床医学等各个领域的应用均非常广泛。

第二节　色谱技术

色谱技术又称层析法（chromatography），俄国植物学家 Michael Tswett 于 1906 年首先创建，当时他用装填有白垩粒子吸附剂的柱子来分离植物叶子色素，各种色素以不同的速率通过柱子时彼此分开，形成易于区分的色素带，由此得名。后来不仅用于分离有色物质而且在多数情况下用来分离无色物质。色谱法由子分离效率高，操作简单等优点而被广泛应用。1941 年 Martin 和 Synge 发现了液－液（分配）色谱［Liquid－Lipuid（partition）Chromatography，LIC］。该法用覆盖于吸附剂表面的并与流动相不混溶的固定液来代替以前仅有的固体吸附剂，使组分按照其溶解度在两相之间分配。在使用柱色谱的早期年代，可靠地鉴定小量的被分离物质是困难的，所以研究发展了纸色谱法（Paper Chromatography，PC）。在这种“平面”的技术中，分离主要是通过滤纸上的分配来实现的。然后由于充分考虑了平面色谱法的优点而发展了薄层色谱法（Thin－Layer Chromatography，TLC），在这种方法中，分离系在涂布于玻璃板或某些

坚硬材料上的薄层吸附剂上进行。气相色谱法是 Martin 和 James 于 1952 年首先描述的，它特别适用于气体混合物或挥发性液体和固体，其特点是分辨率高、分析迅速和检测灵敏等。近年来，因为新型液相色谱仪和新型柱填料的发展以及对色谱理论的更深入了解，又重新引起对密闭柱液相色谱法的兴趣。高效液相色谱（High－Performance Liquid Chromatography，简称 HPLC）迅速成为与气相色谱一样广泛使用的方法，对于迅速分离非挥发性的或热不稳定的试样来说，高效液相色谱常常是更可取的。

一、色谱技术分类

（一）按两相所处的状态分类

液体作为流动相，称为“液相色谱”（liquid chromatography）；用气体作为流动相，称为“气相色谱”（gas chromatography）。固定相也有两种状态，以固体吸附剂作为固定相和以附载在固体上的液体作为固定相，所以层析法按两相所处的状态可分为：液－固色谱（liquid－solid chromatography）、液－液色谱（liquid－liquid chromatography）气－固色谱（gas－solid chromatography）、气－液色谱（gas－liquid chromatography）。

（二）按层析过程的机理分类

吸附层析（adsorption chromatography）是利用吸附剂表面对不同组分吸附性能的差异，达到分离鉴定的目的。分配层析（partition chromatography）是利用不同组分在流动相和固定相之间的分配系数（或溶解度）不同，而使之分离的方法。离子交换层析（ion－exchange chromatography）是利用不同组分对离子交换剂亲和力的不同，而进行分离的方法。凝胶层析（gelchromatography）是利用某些凝胶对于不同组分因分子大小不同而阻滞作用不同的差异，进行分离的技术。

（三）按操作形式不同分类

柱层析（column chromatography）是将固定相装于柱内，使样品沿一个方向移动而达到分离目的的方法。纸层析（paper chromatography）是用滤纸作液体的载体（担体 support），点样后，用流动相展开，以达到分离鉴定的目的。薄层层析（thin layer chromatography）是将适当粒度的吸附剂铺成薄层，以与纸层析类似的方法进行物质的分离和鉴定。

二、离子交换层析

Sober 和 Peterson 于 1956 年首次将离子交换基团结合到纤维素上，制成了离子交换纤维素，成功地应用于蛋白质的分离。离子交换基团不但可结合到纤维上，还可结合到交联葡聚糖（sephadex）和琼脂糖凝胶（sepharose）上。近年来离子交换层析技术已经广泛应用于蛋白质、酶、核酸、肽、寡核苷酸、病毒、噬菌体和多糖的分离和纯化。它们的优点是：①具有开放性支持骨架，大分子可以自由进入和迅速扩散，故吸附

容量大；②有亲水性，对大分子的吸附不大牢固，用温和条件便可以洗脱，不致引起蛋白质变性或酶的失活；③多孔性，表面积大、交换容量大、回收率高，可用于分离和制备。

（一）基本理论

离子交换作用是指在固定相和流动相之间发生的可逆的离子交换反应。离子交换剂通常是一种不溶性高分子化合物，如树脂，纤维素，葡聚糖等，它的分子中引入了可解离的活性基团，这些基团在水溶液中能与溶液中的其他阳离子或阴离子起交换作用。虽然交换反应都是平衡反应，但在层析柱上进行时，由于连续添加新的交换溶液，平衡不断按正方向进行，因此可以把离子交换剂上的离子全部洗脱下来，同理，当一定量的溶液通过交换柱时，由于溶液中的离子不断被交换而浓度逐渐减少，因此也可以全部被交换并吸附在树脂上。如果有两种以上的成分被交换吸着在离子交换剂上，用洗脱液洗脱时，被洗脱的能力则决定于各自洗脱反应的平衡常数。蛋白质的离子交换过程有两个阶段—吸附和解吸附。吸附在离子交换剂上的蛋白质可以通过改变 pH 使吸附的蛋白质失去电荷而达到解离，但更多的是通过增加离子强度，使加入的离子与蛋白质竞争离子交换剂上的电荷位置，使吸附的蛋白质与离子交换剂解离。不同蛋白质与离子交换剂之间形成的键数目不同，即亲和力大小有差异，因此只要选择适当的洗脱条件便可将混合物中的组分逐个洗脱下来，达到分离纯化的目的。离子交换剂还可以经再生处理后反复利用。

（二）离子交换剂的分型

离子交换剂分为两大类，即阳离子交换剂和阴离子交换剂。各类交换剂根据其解离性大小，还可分为强、弱两种。

1. 阳离子交换剂　阳离子交换剂中的可解离基团是磺酸（$-SO_3H$）、磷酸（$-PO_3H_2$）、羧酸（$-COOH$）和酚羟基（$-OH$）等酸性基团。某些交换剂在交换时反应如下：

$$\text{强酸性：}R-SO_3-H+Na^+\rightleftharpoons R-SO_3-Na+H^+$$

$$\text{弱酸性：}R-COOH+Na^+\rightleftharpoons R-COONa+H^+$$

2. 阴离子交换剂　阴离子交换剂中的可解离基团是伯胺（$-NH_2$）、仲胺（$-NHCH_3$）、叔胺［$-N(CH_3)_2$］和季胺［$-N(CH_3)_3$］$^+$等碱性基团。某些交换反应如下：

$$\text{强碱性：}R-N(CH_3)_3OH+Cl^-\rightleftharpoons R-N(CH_3)_2Cl+OH^-$$

$$\text{弱碱性：}R-NH_3(OH)+Cl^-\rightleftharpoons R-NH_3Cl+OH^-$$

离子交换剂根据化学性质分为三类：①纤维素离子交换剂：离子交换纤维素分子上

的羟基被离子基团取代，纤维素离子交换剂对蛋白质的交换容量较其他离子交换剂大，结合力弱，洗脱条件缓和，因而是分离纯化生物活性大分子的理想的工具。阳离子交换剂有羟甲基纤维素（CM—纤维素），阴离子交换剂有氯代三乙胺纤维素（DEAE—纤维素）；②交联葡聚糖离子交换剂：是将交换基团连接到交联葡聚糖上制成的一类交换剂，因而既具有离子交换作用，又具有分子筛效应，是一类广泛应用的色谱分离物质。其非特异性吸附少，不易使蛋白质变性，能引入大量活性基团而不破坏骨架，故交换容量很高。交换容量是指单位重量交换剂中可解离基团的数量，是离子交换剂的特性，与操作条件无关，常用单位重量或单位容积交换剂的能交换的毫克当量离子来表示；③树脂离子交换剂：由苯乙烯和二乙烯苯的共聚物作为骨架，再引入所需的酸性基团或碱性基团，例如聚苯乙烯磺化型离子交换树脂是由苯乙烯与二乙烯苯共聚和再磺化引入磺酸基合成。磺酸根连在树脂上，氢离子与磺酸根的负电荷互相平衡，可与外来离子相互交换。

（三）离子交换层析条件的选择

交换容量是提供交换离子的量度，是选用交换剂数量的重要参考依据。选用何种交换剂还要根据被分离的物质所带电荷、分子大小，及是否耐酸碱等对离子交换剂进行选择。被分离的物质带正电荷，应采用阳离子交换剂，反之则用阴离子交换剂。被分离的物质若为两性电解质，则应考虑在它稳定的 pH 范围内带何种电荷来选择交换剂。对强、中、弱性的选择亦很重要。强性交换剂应用范围广，但选择性低，比如简单的、复杂的、无机的及有机的阳离子都可与强酸性阳离子交换。反之，弱性的交换剂选择性较高，对被分离的物质的活性及稳定性的影响较小。对生物大分子的分离一般选用纤维素和葡聚糖凝胶，其吸附和洗脱条件都比较缓和。

缓冲液的选择原则是阳离子交换剂用阴离子缓冲液，阴离子交换剂用阳离子缓冲液，否则缓冲离子要参与离子交换反应，影响溶液 pH 的稳定。缓冲液的 p H 值和离子强度的选择原则是开始加样时的条件尽可能接近洗脱的条件，对于蛋白质来说一般在其等电点附近选择 pH 值。洗脱液的流速一般来说慢比快好，但太慢有分离时间长、色带扩散等不利因素，因而选择流速有很大的经验性。

三、凝胶过滤层析

凝胶过滤层析技术是 20 世纪 60 年代初发展起来的一种快速而又简单的分离分析技术，由于设备简单、操作方便，不需要有机溶剂，对高分子物质有很高的分离效果。目前已经被生物化学、分子生物学、生物工程学、分子免疫学以及医学等有关领域广泛采用，不但应用于科学实验研究，而且已经大规模地用于工业生产。

（一）基本原理

凝胶过滤层析是使混合物流动相经过固定相凝胶的层析柱时，混合物中各组分按其分子大小不同而分离的技术。固定相凝胶是一种不带电荷的具有三维空间的多孔网状结构的物质，小分子物质可以进入网孔，而大分子物质则被排阻于凝胶颗粒之外，具有分

子筛性质，因而凝胶过滤层析又称为分子筛层析、分子排阻层析等。当混合物样品加入层析柱中，这些物质随流动相地流动而移动，分子量大的物质不能进入凝胶微孔，不受固定相的阻滞，只沿凝胶颗粒的间隙，随流动相向下移动，因此流程短，移动速度快，先流出层析柱；小分子量的物质，可以进入凝胶网孔，阻滞作用大，流程长，移动速度慢，后流出层析柱。分子量大小不同的各组分依据流出层析柱的先后得以分离（见图 8－4）。这种现象叫分子筛效应。

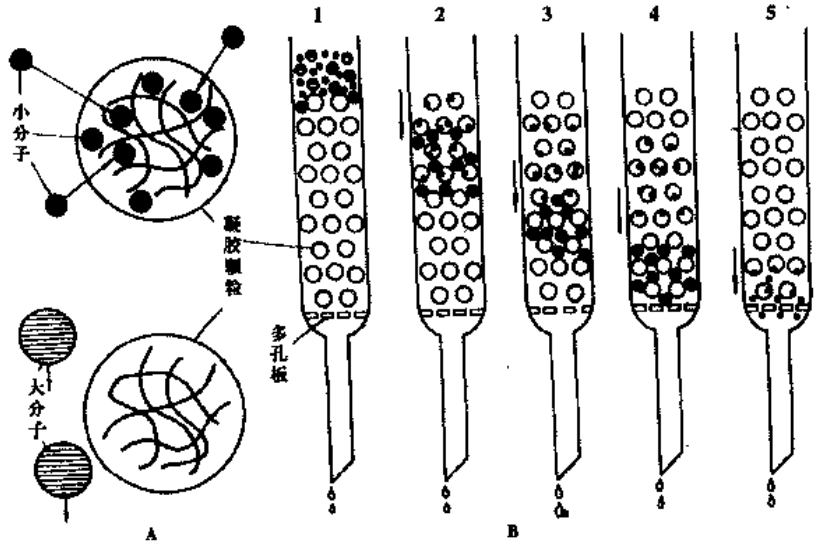


图 8－4　凝胶层析的原理

为进一步说明凝胶色谱的原理，将凝胶装柱后的柱床体积称为总体积（Vt）；色谱柱凝胶颗粒间隙的水相体积称为外体积（outer volume），又称“外水体积”（Vo）；凝胶颗粒内部的水相体积称为内体积（inner volume），又称“内水体积”（Vi）；凝胶支持物的体积简写为 Vg。总体积实际是由 Vo，Vi 和 Vg 三部分组成，即 $Vt=Vo+Vi+Vg$。被分离的物质从加样到组分最大浓度峰出现时所需洗脱液的体积称为洗脱体积（elution volume，Ve）。凝胶总体积可通过测定柱床内径计算出来：$Vt=1/4\pi R^2h$（R－柱内径；h－柱的高度）当柱床装好后，即可加样品，并开始洗脱，分段收集洗脱液，测定其所含组分的浓度，以洗脱液体积为横坐标，组分浓度为纵坐标作图，得洗脱曲线（见图 8－5）。

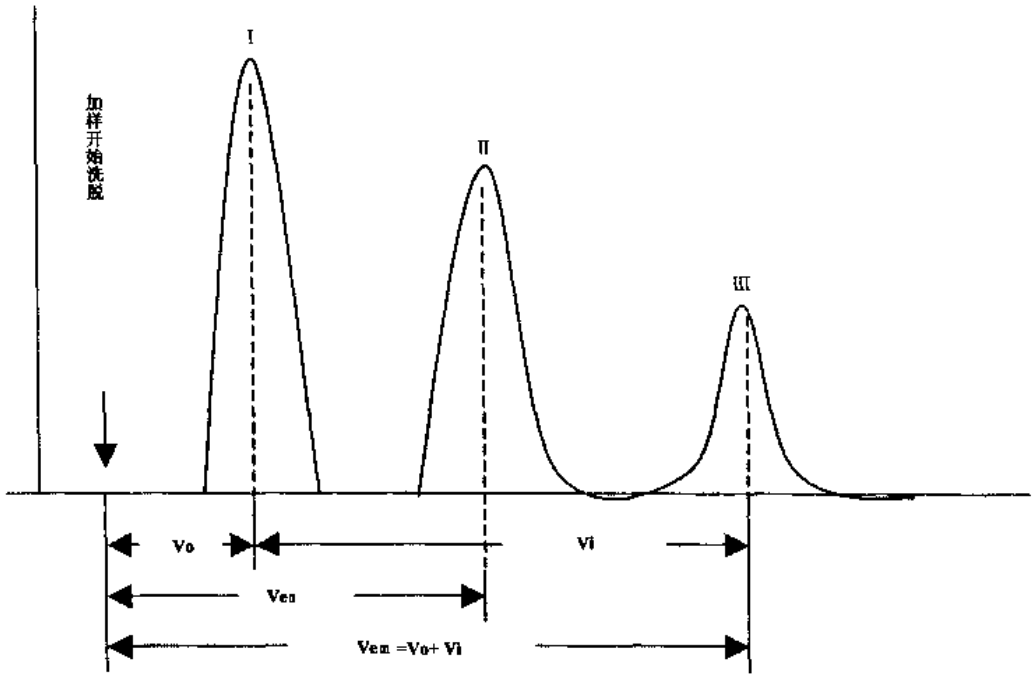


图 8—5 洗脱曲线

$Ve_{Ⅱ}$，$Ve_{Ⅲ}$分别代表组分Ⅱ和Ⅲ的洗脱体积；峰Ⅰ表示大分子完全排阻，峰Ⅱ表示中等大小的分子部分进入凝胶孔隙内，峰Ⅲ表示小分子完全进入凝胶微孔内。

在凝胶层析中分离混合样品时会出现三个高峰，峰Ⅰ是分子很大，完全不能进入凝胶的任何内孔隙的组分；峰Ⅱ是分子大小适中，能进入凝胶的内孔隙中孔径大小相应的组分，峰Ⅲ是分子很小，能进入分子筛全部内孔隙的组分。凝胶层析实际也是一种分配层析，为精确衡量混合物中某一待分离成分在凝胶柱中的洗脱行为，常采用溶质的分配系数 K_d来度量，K_d是溶质分子大小的一个函数。每个组分在两相间有一特定的分配系数，可表达如下：

$$K_d = (Ve - Vo) / Vi$$

K_d 具有如下意义：当某成分的 $K_d = 0$ 时，则 $Ve = Vo$，表示该成分完全被排阻于凝胶颗粒之外，最先被洗脱。K_d 值越接近 0，表示分子越大，洗脱越快。当某成分的 $K_d = 1$ 时，或 $Vi = Ve - Vo$，表示此成分完全不被排阻，可自由扩散入颗粒内部，因此洗脱时，最后流出柱外。K_d 值越接近 1，表示分子越小，洗脱越慢。处于上述两种极端之间的成分，其 K_d 值介于 0—1 之间，其分子量在最大和最小之间。$K_d > 1$ 的物质，表示凝胶对该物质有吸附作用，或形成氢键和离子交换等作用，使阻滞作用增大，此种情况不多见。

（二）凝胶的种类和性质

1. 交联葡聚糖凝胶（Sephadex） 这类凝胶最著名的是瑞典生产的商品名为 Seph-

adex，不同规格型号的葡聚糖用英文字母 G 表示，G 后面的阿拉伯数为凝胶得水值的 10 倍。例如，G－25 为每克凝胶膨胀时吸水 2.5 克，G－200 为每克干胶吸水 20 克。交联葡聚糖凝胶的种类有 G－10，G－15，G－25，G－50，G－75，G－100，G－150，和 G－200。因此，“G”反映凝胶的交联程度，膨胀程度及分离范围。Sephadex LH－20，是 Sephadex G－25 的羧丙基衍生物，能溶于水及亲脂溶剂，用于分离不溶于水的物质。

2. 琼脂糖凝胶　商品名很多，常见的有，Sepharose（瑞典，pharmacia），Bio－Gel－A（美国 Bio－Rad）等。琼脂糖凝胶是依靠糖链之间的次级键如氢键来维持网状结构，网状结构的疏密依靠琼脂糖的浓度。一般情况下，它的结构是稳定的，可以在许多条件下使用（如水，pH4－9 范围内的盐溶液）。琼脂糖凝胶在 40℃以上开始融化，也不能高压消毒，可用化学灭菌法处理。

3. 聚丙烯酰胺凝胶　是一种人工合成凝胶，是以丙烯酰胺为单位，由甲叉双丙烯酰胺交联成的，经干燥粉碎或加工制成粒状，控制交联剂的用量可制成各种型号的凝胶。交联剂越多，孔隙越小。聚丙烯酰胺凝胶的商品为生物胶－P（Bio－Gel P），由美国 Bio－Rad 厂生产，型号很多，从 P－2 至 P－300 共 10 种，P 后面的数字再乘 1000 就相当于该凝胶的排阻限度。

4. 聚苯乙烯凝胶　商品名为 Styrogel，具有大网孔结构，可用于分离分子量 1600 到 40 000 000 的生物大分子，适用于有机多聚物，分子量测定和脂溶性天然物的分级，凝胶机械强度好，洗脱剂可用甲基亚砜。

四、分配层析

在分配层析（也称液－液色谱法）中，溶质分于在两种不相混溶的液相即固定相和流动相之间按照它们的相对溶解度进行分配。固定相均匀地覆盖于惰性载体—多孔的或非多孔的固体细粒或多孔纸上（纸上色谱）。为避免两相的混合，两种分配液体在极性上必须显著不同。若固定相是极性的（例如乙二醇），流动相是非极性的（例如乙烷），那么极性组分将较强烈的被保留，通常称这种方法为正相液－液层析。另一方面，若固定相是非极性的（例如癸烷），流动相是极性的（例如水），则极性组分易分配于流动相，从而洗脱得较快。后一种方法（它有相反的极性）称作为反相液－液层析。由于溶解度差别的细微效应，所以液—液层析很适于分离同系物的同分异构体。在液—液分配层析中，虽然有较好的分离效果，但分离过程中由于机械吸附在载体上固定液的流失，使柱效和分离选择性下降，目前固定相几乎都被化学键合在载体物质上，而不是机械覆盖在它的表面。这种色谱法称作键合相色谱法（Bonded－Phase Chromatography，BPC）。这种方法的机理尚不清楚，可能是分配机理，也可能是吸附机理，视实验条件而定。化学键合固定相对各种极性溶剂都有良好的化学稳定性和热稳定性，因而由它制备的层析柱的柱效高，使用寿命长、重现性好，选择性好。高效液相色谱法中，键合相色谱法的应用远远超过所有其他模式。根据固定相的极性键或非极性键又可分为正反两相。

正相键合相色谱中使用的是极性键合固定相。它是将多孔（成薄壳）微粒硅胶，经

活化处理制成表面含有大量硅羟基的载体后，再与含有胺基（$-NH_2$）、氰基（$-CN$）、醚基（$-O-$）的硅烷化试剂反应，生成表面具有胺基、氰基、醚基的极性固定相。溶质在此类固定相上的分离机理属于分配色谱：

$$SiO_2-R-NH_2 \cdot M + x \cdot M \longrightarrow SiO_2-R-NH_2 \cdot x + 2M$$

SiO_2-R-NH_2 为胺基键合相；M 为溶剂分子；x 为样品组分分子；$SiO_2-R-NH_2 \cdot M$ 为溶剂化的键合相分子；$x \cdot M$ 为溶剂化的组分分子。

在反相键合相色谱法中使用的是非极性键合固定相。它是将多孔（或薄壳）微粒硅胶载体，经活化处理后与含烷基链（C_4，C_8，C_{18}）或苯基的硅烷化试剂反应，生成表面具有烷基或苯基的非极性固定相。其分配机理有两种论点：①分配层析的作用机制：假设在由水和有机溶剂组成的混合溶剂流动相中，极性弱的有机溶剂分子中的烷基官能团（R—）会被吸附在非极性固定相表面的烷基基团上，而溶质分子在流动相中被溶剂化，并与吸附在固定相表面上的弱极性溶剂分子进行置换，由于固定相的极性小于流动相，非极性组分易与固定相结合，即与极性组分分开。从而构成溶质在固定相和流动相中的分配平衡。其机理同正相色谱相似。②吸附色谱的作用机制：溶质在固定相上的保留是疏溶剂作用的结果。根据疏溶剂理论，当溶质分子进入极性流动相后，即占据流动相中相应的空间，而排挤一部分溶剂分子；当溶质分子被流动相推动与固定相接触时，溶质分子的非极性部分（或非极性分子）会将非极性固定相上附着的溶剂膜排挤开，面直接与非极性固定相上的烷基官能团相结合（吸附）形成缔合络合物，构成单分子吸附层。这种疏溶剂的斥力作用是可逆的，当流动相极性减少时，这种疏溶剂斥力下降，会发生解缔，并将溶质分子释放而被洗脱下来。

五、亲和层析

亲和层析技术在 20 世纪 60 年代以后获得了迅速发展，它可在实验室和工业制备规模上，用于分离、纯化具有不同生物活性的物质。其突出的优点是可对天然生物活性物质进行高特异性的分离和纯化。

（一）基本原理

在生物体内，许多大分子具有与某些相对应的专一分子可逆结合的特性。例如抗原和抗体、酶和底物及辅酶、激素和受体、RNA 和其互补的 DNA 等，都具有这种特性。生物分子之间这种特异的结合能力称为亲和力。如果将上述例子中的相互作用体系中的一方连接到基质上，使之固定化，当另一方随着流动相流动时即可与之结合面与其他分子分离，再经过解离洗脱即可获得纯化物质。这种根据生物分子间亲和吸附和解离的原理，建立起来的层析法称为亲和层析技术。亲和层析中两个进行专一结合的分子互称对方为配基。如抗原和抗体，抗原可认为是抗体的配基，反之抗体也可认为是抗原的配基。将一个水溶性配基在不伤害其生物学功能的情况下与水不溶性载体结合称为配基的固相化。

（二）亲和色谱法的基本过程

1. 配基固相化　将与纯化对象有专一结合作用的物质，连接在水不溶性载体上，制成亲和吸附剂后装柱（称亲和性）。

2. 亲和吸附　将含有纯化对象的混合物通过亲和柱，纯化对象吸附在柱上，其他物质流出色谱柱。

3. 解吸附　用某种缓冲液或溶液通过亲和柱，把吸附在亲和柱上的欲纯化物质洗脱出来。

（三）固定相

亲和层析固定相是由基体（Matrix）、间隔臂（Spacer-Arm）和配位体（Ligand）三部分构成，以下分别介绍各部分的组成。

1. 基体　亲和色谱中使用的基体材料可分为天然有机高聚物、合成有机聚合物和无机载体材料三类。此外基体在偶联间隔臂之前还需进行活化预处理。作为亲和色谱的基体材料应具备以下条件：①具有一定的刚性、柔性和机械强度；②具有在一定 pH 值范围的化学稳定性，不溶于普通试剂；②表面具有一定的活性官能团，如羟基、氨基、醚基等，可被适当的试剂活化。当使用基体材料时，还应了解它们的基本物理性质，如粒度、表面积、相对密度、孔径和孔容等必要的参数。

2. 间隔臂　在亲和色谱固定相中，需通过间隔臂将配位体连接在基体上，由于间隔臂占据一定的机动空间，当配位体与被测定的生物分于（尤其是生物大分子）产生亲和作用时，有利于克服存在的空间阻碍作用。当粒度小、孔径小的基体连接小分子配位体时，或当大分于配位体比较密集地分布在基体上时，其与被测定生物分于的亲和作用，会受到空间阻碍作用的影响。为了克服空间阻碍作用，并实现配位体与被测定生物分于间最佳的相互作用。必须解决好作间隔臂的化合物的选择和间隔臂长度的选择。

3. 配位体　在亲和色谱固定相上键联的配位体，可为染料配位体、定位金属离子配位体、包合配合物配位体、生物特效配位体、电荷转移配位体和共价配位体。

（四）流动相

在亲和色谱分析中，分离、纯化的对象皆为氨基酸、多肽、蛋白质、核苷、核苷酸、寡聚和多聚核苷酸、核糖核酸以及酶、辅酶、寡糖、多糖等生物分子，其中大多数为极性化合物，不少还具有生物活性。因此，当从固定相上将它们洗脱下来时，需使用 pH 值接近于中性的稀缓冲溶液，在比较温和的洗脱条件下洗脱，保持其生物活性。

第三节　高效液相色谱

高效液相色谱法（high performance liquid chromatograph，HPLC）为一高效、快速的分离分析技术，具有灵敏度高、选择性好的特点，能解决医药分析领域中的诸多复杂问题。HPLC 法具有同时分离和分析的功能，对于体内药物分析和内源性物质的分

析尤其重要，色谱分离后排除了干扰，大大提高了分析测定的选择性和准确性。HPLC法的分离功能还广泛用于药物的纯化和制备，如用制备规模的色谱柱分离天然药物的有效成分，或制备手性药物的单一对映体。由于使用各种高灵敏度的检测器，再结合许多衍生化技术及样品富集技术，HPLC法对许多药物的最低检测限都能达到pg级或更低水平，非常适合于一些微量成分甚至痕量成分（如杂质或代谢产物）的分析。HPLC法一般在数分钟至数十分钟内即可完成一个样品分析，而且往往能够实现多组分的同时测定，这一快速的特点使HPLC法广泛应用于药物合成的各步反应的监控和临床治疗药物的监测。目前HPLC法技术发展迅速和日趋完善，在蛋白质等大分子物质的分离、纯化和分析中发挥着重要的作用。本节对HPLC法的原理、基本流程及应用作一概要介绍。

一、高效液相色谱基本理论

（一）基本理论

1. 塔板理论的基本假设　色谱过程是物质在相对运动着的两相间连续分配平衡的过程。若两组分的分配系数不等，则被流动相携带移动的速度不一样，使组分达到分离。塔板理论是Martin和Synger首先提出的色谱热力学平衡理论。将色谱柱看成一个分馏塔，把组分在色谱柱内的分馏过程看成在分馏塔中的分馏过程，即把组分的色谱过程分解成在塔板间隔内的分配平衡过程。塔板理论的基本假设为：①色谱柱中存在着许多塔板，组分在塔板间隔内很快达到分配平衡。塔板间隔称为塔板高度（H），组分在塔板高度内完全服从分配定律；②样品的各组分都加在第0号塔板上，样品沿色谱柱纵向扩散忽略不计；③流动相是间歇式进入色谱柱，每次进入一个塔板体积；④在所有塔板上分配系数相等，与组分的进样量无关。

塔板理论的基本假设与实际色谱过程是不同的，整个色谱过程是一个动态的过程，在色谱柱内很难达到分配平衡，而且样品在色谱柱内纵向扩散是不能避免的。但是，塔板理论能解释色谱柱流出的曲线形状呈正态分布，并导出了理论塔板公式用于评价色谱柱柱效。

2. 液相色谱速率理论　液相色谱速率理论方程式是在气相色谱速率理论方程式即Van Deemeter方程式的基础上建立的，Giddings和Snyder等人根据液体和气体的性质差异，提出了液相色谱方程式：

$$H = A+B/u+C_{m}u+C_{sm}u+C_{s}u$$

式中A为涡流扩散项，B/u为纵向扩散项，$C_{m}u$为流动相传质阻抗项，$C_{sm}u$为静态流动相传质阻抗项，$C_{s}u$为固定相传质阻抗项，u为流动相线速度。

涡流扩散项（eddy diffusion）：涡流扩散项$A=2\lambda dp$，dp为色谱柱填料的直径；λ为填充不规则因子，填充越不均匀λ越大。因此应采用小粒度，窄粒度分布及球形或近球形固定相。固定相的直径一般3～10μm，粒度分布RSD≤5%。纵向扩散系数$B=$

$2\gamma D_m$，而 $D_m \propto T/\eta$，在 HPLC 中流动相为液体，其黏度（η）大，柱温（T）低一般在室温，因此 B 很小，若 u>1cm/s，则纵向扩散可以忽略不计。C_m 为流动相传质阻抗系数，C_{sm} 为静态流动相传质阻抗系数。流动相传质阻抗是由于处于一个流路中心和流路边沿的分子的迁移速度不同，而产生峰展宽。静态流动相传质阻抗是分子进入固定相孔穴内的静止流动相中，回到流路中而引起峰展宽。C_s 为固定相传质阻抗系数，在分配色谱中 C_s 与固定液厚度的平方成正比，在吸附色谱中 C_s 与吸附或解吸附速度成反比。只有在厚涂层固定液或深孔离子交换树脂或解吸速度慢的吸附色谱中 C_s 才有明显影响。用单分子层的化学键固定相时 C_s 可以忽略。

（二）基本概念

色谱流出曲线（elution profile）：样品被流动相冲洗，通过色谱柱，进入检测器所得到的信号—洗脱时间曲线称为色谱流出曲线。保留时间（retention time，t_R）：从进样开始到某个组分的色谱峰顶的时间间隔，即从进样到柱后某组分出现浓度最大值的时间间隔。标准偏差（standard deviation，σ）：讨论标准正态分布曲线时，将 X＝±1 时的峰宽之半称为标准偏差。即正常峰标准偏差为峰高 0.607 倍处的峰宽之半。标准偏差的大小说明组分在流出色谱柱过程中的分散程度。σ 小，分散程度小、极点浓度高、峰形瘦、柱效高。反之 σ 大，峰形胖、柱效低。半峰宽（peak width at half－height，$W_{h/2}$）：色谱峰峰高之一半的峰宽称为半峰宽。半峰宽与标准偏差有以下关系：$W_{h/2}=2.355\sigma$。

理论塔板数（theoretical plate number，n）：理论塔板数可以由色谱峰的保留时间和区域宽度计算：

$$n=\left(\frac{t_R}{\sigma}\right) \quad \text{也可写成} \quad n=5.54\left(\frac{t_R}{W_{h/2}}\right)^2=16\left(\frac{t_R}{W}\right)^2$$

用半峰宽计算理论塔板数是最常用的方法。组分的保留时间越长，标准偏差、半峰宽越小，理论塔板数就越大，柱效越高。

分配系数（distribution coefficient，K）：在温度一定时，达到分配平衡后，组分在固定相中的浓度（C_S）与在流动相中的浓度（C_M）之比称为分配系数。

$$K=\frac{C_S}{C_M}$$

在不同分离机制的色谱法中 K 有不同的概念：液－液分配色谱法为狭义分配系数，吸附色谱法为吸附系数，离子交换色谱法为选择性系数，凝胶色谱法为渗透系数，这些分配系数的物理意义虽然各异，但一般情况皆可用狭义的分配系数来描述。在流动相、固定相、温度一定，固定相中的浓度与流动相中的浓度较小时，分配系数只取决于组分的性质，而与浓度无关。这只是理想状态下的色谱条件，在这种条件下，得到的色谱峰

为正常峰；而在许多情况下，随着浓度的增大，分配系数 K 减小，这时色谱峰为拖尾峰；有时随着溶质浓度增大，分配系数 K 也增大，这时色谱峰为前沿峰。因此，只有尽可能减少进样量，使组分在柱内浓度降低，分配系数 K 恒定时，才能获得正常峰。

分离度（resolution，R）：相邻两峰分开的距离与平均峰宽的比值称为分离度

$$R=\frac{(t_{R2}-t_{R1})}{(W_1+W_2)/2}$$

二、高效液相色谱固定相、流动相

（一）固定相的类型和特点

1. 薄壳型　一般在坚实无孔的玻璃核心上包裹一层数微米厚的多孔硅胶，直径约25～70 微米。由于这种填料的颗粒表面薄而均匀，没有深孔，因此促进了样品组分与固定相之间有效的相互作用，使样品组分在微孔中的传质速度加快。减少了谱带扩张，有利提高柱效。并且这种固定相装填容易，渗透性好，有利快速分析。但其表面积小，柱容量有限。

2. 全多孔微球型　是指填料颗粒本身是多孔性物质的固定相。为了加快传质过程，减少谱带扩张，这类固定相一般都制成 5～10 微米的全多孔微球型。由于颗粒小，填料微孔浅，即可改善传质，又可提高柱效。这种固定相柱容量大，分离效果好。但装柱技术要求高，渗透性不如薄壳型好。

3. 化学键合型　它是利用特定的化学反应，把某种有机化合物（固定液）键合到载体表面的特定基团上，如—Si—OH 等，在载体表面形成一牢固的单分子薄层。这种以化学键合的方式代替固定液的机械涂渍，可以克服固定液涂渍不均匀和流失等缺点。化学键合固定相的种类很多，按表面结构分为单分子键合相和聚合键合相两种。按键的类型分为 Si—O—C 键，Si—O—Si—C 键和 Si—O—Si—N 键等几种。按键合固定相的色谱性质分为极性、非极性和离子型等三种。由于键合相键合在载体表面的基团可随意变化，因此，可用于各种类型样品的分离与分析。

（二）流动相

流动相的要求：稳定性好，不与固定相、样品发生化学反应。选择性好，对样品有足够的溶解度。纯度高，黏度小。为了减少流动相中所含微量气体对色谱分析带来的干扰，在应用流动相之前必须对流动相进行脱气处理。常用的脱气方法有：加热脱气法、抽气脱气法、超声波脱气法和吹氮气脱气法。

在色谱分析中，流动相的极性是直接关系着分离效果的重要指标。在液—固吸附色谱中，流动相的极性常用溶剂强度参数 ε° 表示。ε° 的数值越大，流动相的极性越强。在大多数的情况下，单一溶剂很难满足分离复杂样品的需要，这时可使用两种或两种以上不同极性的溶剂，按一定的比例混合，使流动相的极性强度随着组成的变化而变动。在液—液分配色谱中，溶剂的极性常用溶剂强度 δ 表示。δ 值反映了溶剂极性的大小，极

性溶剂 δ 值大，非极性溶剂 δ 值小。同时，也反映了溶剂与样品组分之间的特殊作用。因此，可以通过调整溶剂的 δ 值，以提高溶剂的选择性和分离效能。

色谱分析中流动相的选择回旋余地较大，这是液相色谱的一大特点。但要在种类繁多的溶剂中选择好流动相，并不是容易的事。虽然选择时有“极性相似相溶”的规则可循，但多数仍带有一定的经验性。因为作为液相色谱的流动相，除了要满足上述的一般要求外，还须考虑样品、流动相和固定相的性质以及它们三者之间的相互作用，如酸碱性、电负性，能否形成氢键等。一般情况下，要使样品分离得好，容易洗脱，样品和流动相就应具有化学上的相似性。也就是说，极性大的样品，选用极性大的流动相；极性小的样品，选用极性小的流动相；离子交换色谱则应用 pH 值相近的流动相。对于那些在正相色谱法分离时间较长或难以分离的样品，可改用强极性的流动相和弱极性固定相的反相色谱法进行。有时，为了获得溶剂强度适当的流动相，往往需要经过反复多次的试验，或采用两种以上的混合溶剂作流动相。因此，溶剂系统的选择，不仅取决于待分离的样品化合物，还取决于已经取得的使用分离系统的经验。

三、高效液相色谱仪构成

高效液相色谱仪主要包括输液系统、进样系统、色谱柱系统、检测和数据记录系统。其中输液系统主要为高压输液泵，有的仪器还有梯度洗脱装置；进样系统多为进样阀，较先进的仪器还配有自动进样装置；色谱柱系统除色谱柱外，还包括柱温控制器；数据记录系统，计算机控制系统，进行自动化仪器控制和数据处理。制备型高效液相色谱仪还备有自动馏分收集装置。

（一）输液系统

输液泵的种类很多，按输液性质可分为恒压泵与恒流泵。按结构不同恒流泵又可分为螺旋注射泵、柱塞往复泵和隔膜往复泵。由电动机带动凸轮转动，驱动柱塞在液缸内往复运动。当柱塞被推入液缸时，出口单向阀打开，入口单向阀关闭，流动相从液缸输出，流向色谱柱；当柱塞自液缸内抽出时，流动相自入口单向阀吸入液缸。如此前后往复运动，将流动相不断地输送到色谱柱内。

（二）进样系统

进样器是将样品送入色谱柱的装置。要求进样装置密封性好，重复性好，死体积小，进样时对色谱系统的压力和流量影响小。进样器常用六通进样阀和自动进样系统。六通阀的构造及进样过程：当进样阀处于装样位置时，用微量注射器将样品注入定量管。进样时转动进样手柄至进样位置，定量管内的样品由流动相带入色谱柱。进样体积是由定量管控制的，所以进样量准确，重复性好。为了进样量的准确，用微量注射器吸取样品时，要大于定量管容量的 2～3 倍。如果需要小体积进样，用微量注射器吸取样品时，要小于定量管容量的 1/2 倍。另外，转动进样手柄时不能太慢，更不能停留在中间位置，否则流动相受阻使柱内压力下降，泵内压力增大，造成仪器损坏。

（三）色谱柱

色谱柱的正确使用和维护是十分重要的，如果不注意这一点，一是使柱效降低，二是缩短色谱柱的使用寿命，甚至损坏色谱柱。尽量减少柱压和温度的急剧变化及任何机械震动。色谱柱尽量不要反冲，否则反冲会降低柱效。不要使用 pH3～8 范围以外的流动相，否则容易损坏色谱柱中的固定相。避免将基质复杂的样品，尤其是生物样品注入色谱柱之前，需要对样品进行预处理（分离或过膜），并要求在进样器与色谱柱之间连接与色谱柱固定相相同的保护柱，同时注意经常要更换保护柱。要经常用强溶剂冲洗色谱柱，以除去存留在柱内的杂质。在进行清洗时，对流动相的置换应以相混溶的溶剂逐渐过渡，每种流动相的体积一般 50～70mL。硅胶柱的清洗溶剂及顺序：正己烷、二氯甲烷、甲醇，然后再反顺序清洗。反相柱的清洗溶剂及顺序：水、甲醇、乙腈、氯仿，然后再反顺序清洗。色谱柱长期不用时，柱内要用乙腈或甲醇充满。禁止缓冲溶液留在柱内过夜或长期存放。色谱柱使用一段时间后，柱压可能升高。其原因一种可能是柱进口端的烧结滤片被堵塞，这时应清洗或更换滤片。另一种可能是大分子进入色谱柱柱头被污染，或者是柱头出现窝陷。这时可打开柱接头，取出柱头 1～2mm 深度的填料，再用相同填料将柱头填平并拧紧柱接头。这样色谱柱的柱效能得到改善，但难以恢复早期的柱效。

（四）检测器

检测器是高效液相色谱的主要关键部件之一。它的作用是从色谱柱流出的组分由检测器转变成电信号。检测器按其适用范围可分为专属型和通用型两大类。专属型检测器只能检测某些组分的某一性质，如紫外检测器、荧光检测器，它们只对有紫外吸收或荧光发射的组分有响应；通用型检测器可检测一般物质均有的性质，如示差折光检测器就属于这类检测器。高效液相色谱的检测器灵敏度高、噪音低、线性范围宽、重复性好、使用范围广。

1. 紫外检测器（ultraviolet detector UV） 是高效液相色谱中最常用的检测器，多数紫外检测器检测波长包括可见波长时，又可称为可见—紫外检测器。它灵敏度高，对流速和温度的波动不灵敏。紫外检测器的工作原理是朗伯—比尔（Lambert－Beer）定律，当一束单色光透过流动池时，若流动相不吸收光，则吸光度（A）与吸光组分的浓度（C）和流动池的光径长度（b）成正比。

$$A=Ig\ (I_0/I)\ =Ig\ (I/T)\ =abC$$

式中 I_0 为入射光强度，I 为透射光强度，T 为透光率，a 为吸光系数。

紫外检测器包括固定波长和可变波长检测器及光电二极管阵列检测器。

（1）固定波长检测器：检测器的波长一般为 254nm，低压汞灯为光源，光源单色性好，光强度大，灵敏度高。因此，这种检测器对许多有紫外吸收的化合物，如具有芳环、芳杂环的化合物都能进行检测。共轭性越强的化合物吸光系数越大，检测灵敏度

越高。

（2）可变波长检测器：这是一种最常用的检测器，一般采用氘灯，或氢灯为光源，能够按需要选择组分的最大吸收波长为检测波长。从而提高灵敏度。但是，光源发出的光是通过单色器后照射到样品上，因此单色光强度相对较弱。这类检测器的光路系统和流动池构型与固定波长检测器相似。

（3）光电二极管阵列检测器：这种检测器是20世纪80年代出现的一种光学多通道检测器。在晶体硅上紧密排列一系列光电二极管，每一个二极管相当于一个单色仪的出口狭缝。当复光通过流动池时，被组分选择性吸收后，再进入单色器，照射在二极管阵列装置上，使每个纳米波长的光强变成相应的电信号强度，而获得组分的吸收光谱。经过计算机处理，将每个组分的吸收光谱和样品的色谱图结合在一张三维坐标图上，而获得三维光谱－色谱图。

2. 荧光检测器（fluorophotometric detector）　这种检测器的灵敏度比紫外检测器高，但是只适合于能产生荧光物质的检测。如具有天然荧光的卟啉、生物胺、维生素及甾体化合物，通过荧光衍生化使没有荧光转变成具有荧光物质的化合物都可以用这种检测器来测定。其工作原理是，具有某种特殊结构的化合物，受紫外光激发后，能发射出比激发光源波长更长的光，称为发射光；荧光强度（F）与激发光强度（I_0）及荧光物质浓度（C）之间的关系为：

$$F=2.3QKI_0\varepsilon cl$$

式中Q为量子产率，K为荧光收集效率，ε为摩尔吸光系数，l为光径长度。

一般激发波长（λex）与化合物的最大吸收波长（λmax）相近。选择激发波长时，把发射单色器固定在某一波长处，开大发射单色器的狭缝，改变激发波长进行扫描，得激发光谱，光谱上的峰对应的波长即为激发波长。发射波长（λex）的选择是把激发单色器固定在λex处，改变发射波长进行扫描，得荧光发射光谱，光谱上的峰对应的波长即为发射波长。检测多组分时，可改变发射单色器的波长，以获得较多色谱峰和较高灵敏度的波长为发射波长。

3. 电化学检测器（electrochemical detector，ECD）　包括极谱、库仑和安培检测器等。其中安培检测器应用最广泛，安培检测器的工作原理是在电极间施加一恒定电位，当电活性组分经过电极表面时，发生氧化还原反应，电量（Q）的大小符合法拉第定律$Q=Nfn$。因此反应的电流（I）为

$$I=nF\frac{dN}{Dt}$$

式中n为每摩尔物质在氧化还原过程中转移的电子数，F为法拉第常数，N为物质的摩尔数，t为时间。当流动相流速一定时，dN/dt与组分在流动相中的浓度有关。

安培检测器的灵敏度很高，凡具氧化还原活性的物质都能进行检测，如生物胺、酚

羰基化合物、巯基化合物等。本身没有氧化还原活性的化合物经过衍生化后，也能进行电化学检测。

四、高效液相色谱分析方法

高效液相色谱在医药领域中应用非常广泛。能应用于化学合成药的含量测定和杂质限度检查、制剂分析、药物代谢和动力学研究，还能进行蛋白质成分的分离和制备、纯化。这些分析方法大致可分为定性分析和定量分析。

（一）定性分析

HPLC 定性分析是利用色谱定性参数保留时间对组分进行定性分析，其原理是同一物质在相同色谱条件下保留时间相同。此法适用于已知范围的未知物，将样品中加入某一纯物质，混合进样，对比加入前后的色谱图，看其峰形增多还是增大，峰形增多，则不是一种物质，峰形增大，则是一种物质。为了提高定性的可靠性，还应改变色谱分离条件，作进一步验证。

（二）定量分析

HPLC 定量分析是利用色谱峰峰高和峰面积与样品组分的含量呈正比。常用的定量方法有外标法和内标法。

1. 外标法　用与待测样品同样的标准品作对照，以对照品的量对比待测样品含量的分析方法称为外标法。只要待测样品组分出峰、无干扰、保留时间适当，就可用此法进行定量分析。要求进样量必须准确，由于在 HPLC 中进样常用六通阀进样，误差较小，所以外标法是 HPLC 常用定量分析方法。外标法分为外标一点法及外标工作曲线法。

（1）外标一点法：用一种浓度的对照品溶液对比求算样品含量的方法，称为外标一点法。外标一点法的计算公式如下：

$$C_{样品}=C_{对照}\times A_{样品}/A_{对照}$$

式中 $C_{样品}$ 与 $A_{样品}$ 为样品的浓度与峰面积；$C_{对照}$ 与/$A_{对照}$ 分别为对照品的浓度与峰面积。

（2）外标工作曲线法：用对照品配制系列浓度的对照品溶液，准确进样，测得峰面积（A）或峰高（h），对浓度 C 绘制工作曲线，计算回归方程，用回归方程计算样品溶液的含量。

$$A=a+bc$$

式中 b 与 a 分别为直线的斜率和截距。

2. 内标法　选择适当的物质作为内标物，以待测样品组分和内标物的峰面积比计算样品含量的分析方法称为内标法。此法所选用的内标物应是样品中不含有的物质，其

保留时间与待测样品组分靠近，并且不与其他峰相互干扰。使用内标法可以抵消仪器稳定性差、进样量不够准确等因素带来的定量分析误差。如果在样品预处理前加入内标物，可抵消方法全过程引起的误差。

（1）工作曲线法：本法与外标工作曲线法相似，只是在各种浓度的对照品溶液中加入同量的内标物。分别测定i组分与内标物s的峰面积A，以峰面积Ai/As与Ci绘制工作曲线，计算回归方程。

$$Ai/As=a+bCi$$

将与对照品溶液中相同的内标物加到样品溶液中，分别测定样品中i组分与内标物峰面积，以两者峰面积之比代入回归方程计算出样品中i组分的含量。

（2）内标对比法：此法只配制一个浓度的对照品溶液，然后在样品与对照品溶液中加入同量的内标物，分别进样。按下式计算样品浓度：

$$(Ci)\ 样品=\frac{(Ai/As)\ 样品}{(Ai/As)\ 对照}\times(Ci)\ 对照$$

与外标法相似，只有线性关系截距为零时才可用此法定量。

五、氨基酸、多肽和蛋白质的HPLC分析

（一）氨基酸的HPLC分析

氨基酸是蛋白质构成的基本单位，由于其参与肌体正常代谢和许多生理机能而受到重视。自20世纪80年代以来，氨基酸、多肽和蛋白质类药物在临床上的应用增多处，HPLC在分析这些物质方面也得到广泛的应用。

氨基酸是一类分子中同时具有氨基和羧基的化合物，为两性电解质。动物体内组成蛋白质的氨基酸均为α—氨基酸，约20余种，分为脂肪族氨基酸、芳香族氨基酸及杂环氨基酸。通常多数氨基酸容于极性溶剂，只有苯丙氨酸和酪氨酸难溶于水。由于氨基酸结构及性质既有相似性又有其各自特点，其混合物的分离测定有较大难度，HPLC法在这一方面显示了特有的优越性。但大多数氨基酸无紫外吸收和荧光发射，所以，为提高分析的检测灵敏度通常将氨基酸衍生化，衍生化方式有柱前衍生化和柱后衍生化。由于柱后衍生化所需色谱仪器复杂，操作烦琐，因而应用不广泛。近十年来，随着RP—HPLC及各种柱前衍生化试剂的出现，而使得氨基酸分析的灵敏度有很大的提高，分析方法简便、快速。

例如可用离子交换 HPLC 法测定人体生理氨基酸，其色谱条件是 ①色谱柱：阳离子交换柱(15cm×3mm)；②流动相：A. 缓冲液（89.2g 枸橼酸三锂四水合物、2mol/L 氢氧化锂 125mL、硫代乙二醇 25mL、苯酚 6.26mL 溶子 4.5L 水中。将溶液用盐酸调至 Ph2.27，加水至 5.0L)；B. 缓冲液（取 277.3g 枸橼酸三锂四水合物、5.0gEDTA，其余同流动相 A，调至 pH7.50)；③梯度及流速：见表 8－3；④检测：荧光 340nm/448nm；⑤衍生化。

表 8－3 流动相梯度及流速

时间 min	A（%）	B（%）	流速 mL/min
0	100	0	0.2
35	95	5	0.2
40	95	5	0.2
75	85	15	0.2
80	40	60	0.3
110	35	65	0.3
145	0	100	0.3
195	0	100	0.3

样品测定过程如下：取空腹过夜健康供试者静脉血，每 1mL 血中加 EDTA 三钾 1mg，离心分离血浆，于－80℃保存。取血浆 200μL 于冷的 16%（W/V）磺基水杨酸、400nmol/mLα－氨基乙二酸 50μL 旋涡混合，4℃放置 10min，再旋涡混合，4℃离心 10min，上清液滤过。取滤液 7μL 进样，保留时间：4.34～175.7min 内出峰。除蛋白后人血浆中氨基酸的柱后 OPA 衍生化色谱图见图 8－6。

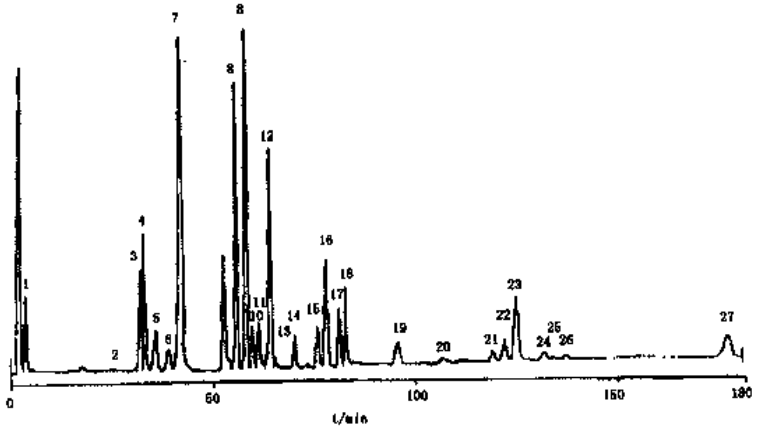


图 8－6 除蛋白人血浆中氨基酸的 OPA 柱后衍生化色谱图

1. 牛磺酸 2. 天门冬氨酸 3. 苏氨酸 4. 丝氨酸 5. 天冬酰胺 6. 谷氨酸 7. 谷氨酰胺 8. 甘氨酸 9. 丙氨酸 10. 瓜氨酸 11. 氨基丁酸 12. 缬氨酸 13. 胱氨酸 14. 蛋氨酸 15. 异亮氨酸 16. 亮氨酸 17. 酪氨酸 18. 苯丙氨酸 19. 色氨酸 20. 氨基乙醇 21. 鸟氨酸 22. 赖氨酸 23. 组氨酸 24. 甲基赖氨酸 25. 3－甲基组氨酸 26. 甲基组氨酸 27. 精氨酸

（二）蛋白质的 HPLC 分析

色谱法用于蛋白质分离已有数十年的历史。所采用的 HPLC 法有反相色谱、离子交换色谱、凝胶过滤色谱、亲和色谱等。反相色谱是基于蛋白质在极性流动相和非极性固定相间相互作用的差别将蛋白质加以分离，流动相多采用低离子强度的酸性水溶液，加入有机溶剂作为有机改性剂，加入三氟醋酸或盐酸作为离子抑制剂。离子交换色谱利用蛋白质分子和离子交换剂之间发生的静电作用使填料表面的可交换基团与带相同电荷

的蛋白质分子发生交换，采用的流动相为盐－水体系，通过调节 pH 梯度或盐离子强度梯度而分离蛋白质。凝胶过滤色谱法按蛋白质分子大小加以分离，填料具有一定孔径，大分子蛋白质不能进入小孔穴先出柱，洗脱液为盐的水溶液。亲和色谱利用蛋白质和固定相表面的某种特异吸附而进行选择性分离，没有这种特异吸附作用的分子不被保留而先出柱，然后改变流动相条件，将保留在柱上的样品组分洗脱。

用于蛋白质分离的 HPLC 系统应具有梯度洗脱功能，检测采用紫外检测或荧光检测，紫外检测波长为 280nm，254nm 或 215nm，蛋白质在 215nm 处吸收较强，故在该波长处检测灵敏度高，但对溶剂要求较高；荧光检测要求多数蛋白质应在检测前进行衍生化反应。

多肽类药物为分子量较小的蛋白质，由于其同时具有氨基酸和蛋白质的共同性质，分子量较小的肽类化合物采用的 HPLC 法有许多与氨基酸较接近，分子量较大的肽类更多使用与蛋白质相同的方法。用于肽类药物测定经常使用的 HPLC 法为 RP－HPLC 法，使用的固定相为十八烷基、苯基和睛基键合硅胶；填料孔径应在 10nm 以上，分子较大的肽则需要更大孔径填料，色谱柱长度多在 15cm 以下；流动相为缓冲溶液或离子对溶液；洗脱方式较多采用梯度洗脱，也可用等溶剂强度洗脱。离子交换色谱在肽类药物的分析中也较为常用。流动相多为与 RP－HPLC 相同的缓冲液；亦可采用梯度洗脱方式。肽类药物 HPLC 的检测多用紫外 200～230，280nm 附近也可使用，但吸收较弱；一些小分子肽类药物，经衍生化以后用荧光检测器检测可提高检测灵敏度。

现以人胃液中的胃蛋白酶的高效离子交换色谱分离为例简单介绍如下。胃蛋白酶（Pepsin）是动物胃液中多种蛋白水解酶的混合物，含有胃蛋白酶、组织蛋白酶、胶原蛋白酶等。由猪胃黏膜中提取的胃蛋白酶为白色或淡黄色粉末，易溶于水。胃蛋白酶的活性依据胃蛋白酶在一定条件下水解蛋白底物程度判断。HPLC 分离胃蛋白酶色谱条件是 ①色谱柱：TSK DEAE（7.5cm × 7.5mm，10μm）；②预柱：TSK DEAE 柱；③流动相：A. 50mmol/L 醋酸钠溶液（pH4.1）；B. 1mol/L 氯化钠；④梯度：0～5min，0%B；5～30min，0%～30%B；30～30.1min，30%～100% B；30.1～40min，100% B；40～42min，100% ～ 0% B；⑤检测：紫外，280nm。样品分离过程是，收集胃液（至 4L）通过 Whatman 113v 滤纸滤过。用 Sartorios 弦向流动超滤仪浓缩至 800mL，用 50mmol/L Ph4.1 醋酸钠溶液 10L 透析 16h，再用 100g 预先用 50mmol/L pH4.1 醋酸钠溶液平衡的 DEAE 纤维素混合，搅拌 1h，静置，弃取上层液。胃蛋白酶 3，5 用含有 0.25mol./L 氯化钠的 50mmol/L pH4.1 醋酸钠溶液

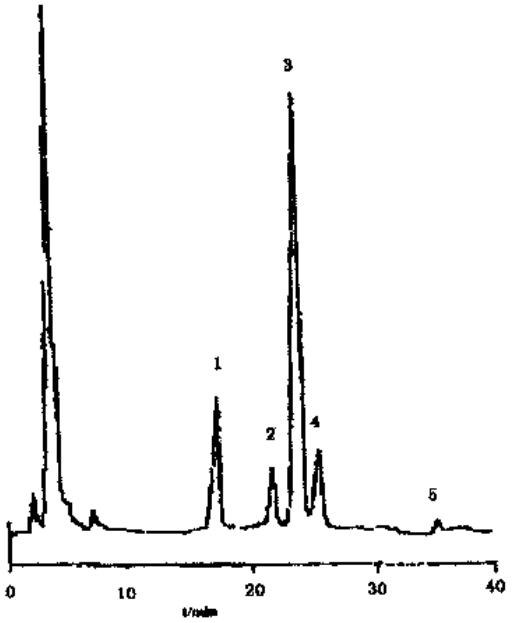


图 8－7　胃液中蛋白酶色谱图

1. 胃蛋白酶 5　2. 胃蛋白酶 3c　3. 胃蛋白酶 3b　4. 胃蛋白酶 3a　5. 胃蛋白酶 1

750mL 洗脱；胃蛋白酶 3，1 用含有 1mol/L 氯化钠的 50mmol/L Ph4.1 醋酸钠溶液 750mL 洗脱，洗脱液 200mL 用超滤法浓缩至 10mL，用 50mmol/L pH4.1 醋酸钠溶液 1L 透析 16h。取经透析的溶液 500μL 进样。分析色谱图见图 8—7。

（胡海燕　刘传华）

第九章　蛋白质结构分析技术

第一节　蛋白质和多肽的氨基酸序列分析

蛋白质的种类繁多，不同的蛋白质行使不同的功能，这是通过其特定的结构来实现的。一级结构决定高级结构，蛋白质的功能归根结底是由一级结构决定的。因此，测定蛋白质的氨基酸序列是进行蛋白质结构或功能研究中不可缺少的部分。目前，进行蛋白质序列测定首先采用化学法或酶法从蛋白质或多肽的 N 端或 C 端进行裂解，将氨基酸残基逐个从蛋白质或多肽的末端切割下来，然后通过在氨基酸残基上衍生一个生色基团，采用薄层层析、高效液相色谱等技术鉴定每次切割下来的氨基酸残基。

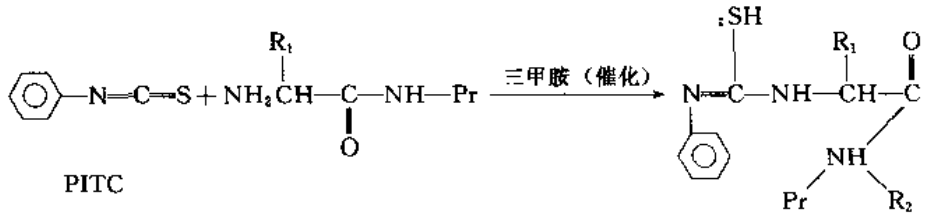
一、N 端序列分析

N 端序列分析的方法很多，最常用的是化学法，有 Edman 降解法、DNS－Edman 降解法、减数 Edman 法和 DABITC/PITC 双偶合法。Sanger 早年使用的 FDNB 法，由于灵敏度低、操作较复杂和不能连续降解等原因，人们一般不再采用。用酶学方法测序虽有一定意义，但在实际应用中尚有许多弊病，因此很少有人用。这里介绍 Edman 化学降解法。

（一）Edman 化学降解法

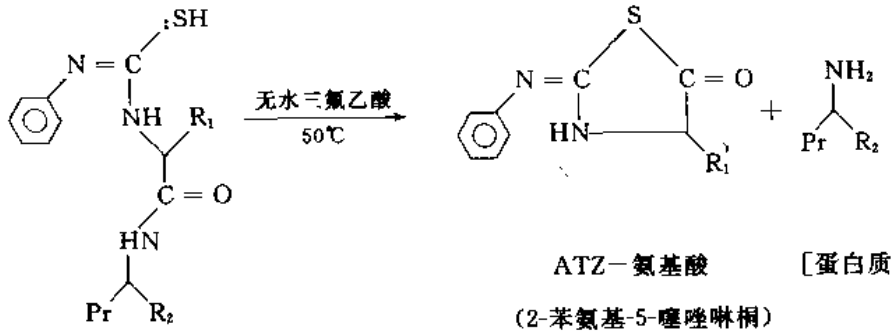
Edman 化学降解法迄今仍然是蛋白质或多肽顺序测定技术中最有效和最基本的方法，是 1950 年由 Edman 首先提出并以他的名字命名的一种化学降解法。降解所用的试剂是异硫氰酸苯酯（PITC），它与蛋白质或多肽的自由 α－氨基作用。降解反应分三步：

1. 偶联反应　PITC 的异硫氰酸酯基与蛋白质或多肽的自由 α－氨基作用，产生苯氨基硫甲酰蛋白质。反应所需条件：pH9.0～9.5（PITC 仅仅与未质子化的氨基反应，pH≥10 时很不稳定）；为保证反应速度，50℃左右较合适，仅需 10 分钟；注意不同的副反应的干扰，尤其是空气中的氧和试剂中含的醛类杂质会引起氧化脱硫产生异氰酸苯酯肽。



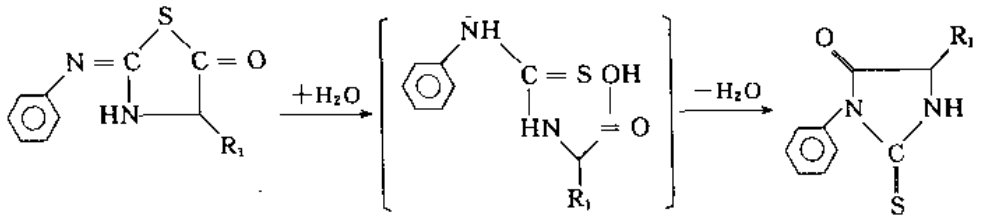
PITC　　PTC—蛋白质
（苯氨基硫甲酰肽）

2. 环化裂解反应　PTC—蛋白质在无水条件下酸裂解，切下反应的那个残基，紧挨的第二个氨基酸残基暴露出自由的 α—氨基，可与 PITC 继续进行偶联反应。



ATZ—氨基酸　　[蛋白质—1]
（2-苯氨基-5-噻唑啉酮）

3. 转化反应　因为 ATZ—氨基酸不稳定，直接鉴定比较困难，需在三氟乙酸水溶液（25%）条件下，转化成稳定的 PTH—氨基酸。



ATZ—氨基酸　　PTC—中间体　　PTH—氨基酸
（苯乙内酰硫脲衍生物）

4. PTH—氨基酸的鉴定　降解产生的 PTH—氨基酸可采用多种方法进行分析鉴定（见表 9—1）。

表 9－1　PTH－氨基酸分离及鉴定方法

方法	前处理	检出物	同时分析	分析时间	操作温度	定量	灵敏度
氨基酸分析	150℃ 5.7N HCl 水解 4 小时	除 PTH－Gln，Asn Trp 外的大多数 PTH－氨基酸	不可	1 小时	高温	能	1nmol/L
薄层层析		所有 PTH－氨基酸	可以	0.5 小时	室温	不能	0.1nmol/L
气相色谱	甲基硅烷化	除 PTH－Arg，Cys 外的大多数氨基酸	不可	25 分钟	高温	能	1nmol/L
高效液相色谱		所有 PTH－氨基酸	不可	30 分钟	室温	能	0.1nmol/L
质谱分析		除 PTH－Arg 外的所有	不可	20 分钟	高温	能	5pmol/L

除表 9－1 所列主要方法外，还有几种，如光谱法，在 269nm 能监测所有的 PTH-氨基酸，在 320nm 监测脱水的 PTH-Ser 和 Thr；高压电泳能监测 PTH-Arg，His，$CySO_3H$；菲醌试剂监测 PTH-Arg；Pauly 试剂监测 PTH-His，Tyr 等。近年来，由于高效液相色谱技术具有分析速度快、灵敏度高、定性定量准确等诸多优点，同时仪器自身结构也日臻完善，多被广泛应用于 PTH-氨基酸的鉴定。通常选用 C_{18} 反相柱进行分离。

（二）蛋白质 N 端测序仪器分析法

自 20 世纪 70 年代初蛋白自动序列仪诞生以来，经过了一系列更新换代，仪器日趋灵敏、快速，下面就其中几种代表性的仪器作一简介。

1. 液相旋转杯序列仪　这是最早的自动化仪器。整个仪器的核心部分是一个反应杯，它由一个马达带动并有调速控制。蛋白质或多肽样品经溶解后，通过一根导管注入反应杯，随着旋转的离心力被均匀地涂在杯壁上，形成一层薄膜，其厚薄可由转速控制。反应的试剂和溶剂分别通过导管进入反应杯，与杯内薄层上的样品的 N 端自由氨基进行 Edman 反应；降解下来的 ATZ 氨基酸衍生物通过另一导管引出并收集，再将 ATZ 氨基酸转换成 PTH 氨基酸后进行鉴定，依次循环分析。此仪器由于消耗的样品量多，目前在蛋白质化学实验室已很少使用

2. 固相序列分析仪　自动固相顺序仪同液相顺序仪一样，都是在电子计算机的控制下工作。由于一些小肽或疏水性多肽，不易吸附在一般支持物上（如旋转杯），只得将它们共价键合到一种惰性支持体即载体上，然后装入一根小口径的反应柱，随后再将反应用的试剂和溶剂分别按程序注入反应柱内进行 Edman 化学降解反应。早先的固相顺序仪只能完成 Edman 反应式中的偶联裂解二步反应，第三步转化反应和最后的 PTH 氨基酸鉴定工作需另外进行。近年来，有的型号的仪器已配有自动转换装置。一般整个转换反应共五步：①从反应柱上裂解下来的 ATZ－氨基酸衍生物引入转换器的反应杯

内；②通 N_2 干燥；③加入可溶性酸到反应杯内，进行转换反应；④再次通 N_2 干燥；⑤加萃取溶剂，将反应杯中转换后的 PTH－氨基酸引出，注入部分收集器的试管中用于鉴定。固相测序的优点是样品共价结合后，不会在有机溶剂洗涤、抽提等过程中丢失，因而可以增加循环次数。其缺点是共价结合不外乎通过氨基或羧基与载体上的活性基团作用，蛋白质和多肽中的谷氨酸、天冬氨酸、赖氨酸等含有除 α－氨基或 α－羧基以外氨基或羧基的氨基酸残基将结合在载体上，不能正确确定。

3. 气相顺序仪　在蛋白质或多肽顺序分析的发展过程中，自动液相顺序仪和固相顺序仪曾起过积极作用，但是随着分子生物学的深入发展，近年来人们深感这两种仪器已不能满足蛋白质微量顺序分析的需要。1981 年 Hewick R. M. 等人研制了一种新型顺序仪，即气相顺序仪，亦称气液固相顺序仪。其特点是弹筒型的反应室代替了液相顺序仪中的旋转杯，反应试剂是在气态下完成蛋白质的 Edman 降解反应。其优点是灵敏度很高，作为顺序测定的蛋白质或多肽量最低在 5pmol 水平；试剂和溶剂消耗很低，大约是液相仪的十分之一；缩短了每步降解循环时间，因此提高了效率；降低了费用。

4. 新型蛋白测序仪　20 世纪 80 年代末又对气相序列仪进行了部分改进，将原来三氟乙酸以气体方式输送改为液相脉冲方式，称为脉冲液相序列仪，以适应不同样品的分析需要。另外，由于载有活性基团的功能性聚亚乙烯基二氟（PVDF）膜的问世，蛋白质和多肽可以共价固定在此类膜上，直接在上述两种序列仪上进行固相序列分析。即经 Edman 降解后的 PTH－氨基酸衍生物可及时直接从转化腔中自动输入高效液相色谱仪中进行 PTH－氨基酸衍生物的定性及定量分析，不仅快速可靠，而且防止了样品的损失。将序列仪与高效液相色谱仪联用，并统一用电脑控制，包括化学反应和分析鉴定以及数据的自动记录和处理。新近推出的“173 毛细管 HPLC 及微量蛋白印迹制备系统”将分离、印迹、电泳、纯化、制备、N-端测序为一体，是一类高自动化的蛋白测序系统。图 9－1 为标准 PTH-氨基酸的色谱分离图，欲在 20 分钟内将各组分较好分离，需选择合适的色谱条件：色谱柱：Lichrosorb RP-18 Hibar 5μm（300×4.6）；流动相：68.5% 0.01mol/L 醋酸钠（pH5.4），31%乙腈，0.5% 1，2－二氯乙烷；柱温：62℃；流速：1.5mL/min。

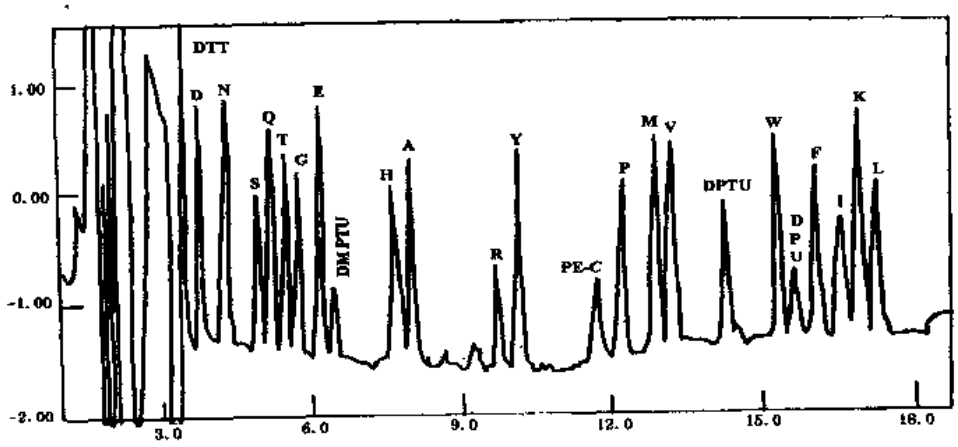


图 9－1　PTH－氨基酸色谱图

（三）N 端测序前样品处理

1. 纯度鉴定　宜采用多种互补有效的手段鉴定样品的纯度，如 SDS－PAGE、反相 HPLC、毛细管电泳、阴离子或阳离子的 FPLC 等。

2. 脱盐　可采用反相 HPLC、凝胶过滤、透析、超滤等，所用试剂、仪器乃至操作规范都必须是测序级的，应尽量避免引入新的杂质。对于普通实验室和蛋白含量少的样品一般采用简便的 ProSpin 装置，用 ProBlott PVDF 和 3000 道尔顿截留过滤膜离心，除去样品中的缓冲盐、去垢剂及其他小分子杂质。然后将插管倒置，用切割器取下吸附有蛋白质的膜片。对于特别稀少的样品可将切下的膜片分割成几部分，分别进行测序、氨基酸组成分析、肽谱分析等工作，其中每部分约需 50pmol。

3. 巯基修饰方法　采用丙烯酰胺（包括还原、烷基化、脱盐 3 步）和 4－乙烯吡啶（还原、烷基化、脱盐）修饰方法。

4. 去除 N 端封闭的基团的方法　若怀疑蛋白质或多肽的 N 端被封闭，可采用：①去除 N－乙酰丝氨酸和 N－乙酰苏氨酸残基；②去除 N－端甲酰甲硫氨酸中的甲酰基；③去除 N－端的焦谷氨酸；④去除 N－乙酰基封闭的氨基酸残基。

二、C 端序列分析

虽然建立在 Edman 化学法上的 N 端测序技术日趋成熟，但仍有不少问题需要借助 C 端序列分析来解决。C 端测序尤其适用于基因重组蛋白是否正确表达的鉴定和大规模生产时的质量控制以及 N 端封闭蛋白的分析和 DNA 探针的设计。

由于选择性对 C 端羧基进行偶联修饰并从 C 端逐一将氨基酸残基切下较 N 端测序困难，在过去的近 20 年里，虽然有些蛋白质化学家致力于 C 端测序研究，但目前仍不能和 N 端测序技术媲美。近几年来，由于对化学法测定 C 端技术进行了一系列改进，已有了自动化 C 端序列仪问世。下面介绍一种自动化 C 端测序方法的原理及应用。

（一）C 端分析原理

基于与 N 端 Edman 降解类似的原理，C 端分析采用化学试剂与蛋白质和多肽的 α－羧基反应，反应后的 C 末端衍生物被切割下来，通过 HPLC 系统进行分离分析。一个完整的 C 端序列分析包括以下几个步骤。

1. 活化　蛋白质和多肽 C 端的自由羧基与乙酸酐反应生成混合酸酐，并进一步环化成咪唑酮，而侧链上的羧基形成的混合酸酐则难以成环。C 末端的咪唑酮在四丁基硫氰酸氨的作用下，转化为乙内酰硫脲（TH）。

2. 酰胺化保护侧链羧基　侧链羧基形成混合酸酐后，与硫氰酸哌啶进一步作用形成酰胺，以保护侧链。

3. 烷基化　由于 C 末端环化成的 TH 衍生物较为稳定而不易被切割，因此产率不高。用溴代甲基萘选择性地烷基化修饰硫原子，形成烷基化－乙内酰硫脲（ATH）衍生物，裂解产率将大大提高。

4. 修饰 Thr 和 Scr 的羟基　蛋白质和多肽中的羟基会干扰测序，因此需要修饰。

在活化过程中，乙酸酐也会与羟基反应将其乙酰化，但反应不完全，在 N－甲基咪唑（NMI）和乙酸酐的共同作用下，Thr 和 Ser 上的羟基基本被乙酰化。

5. 裂解和衍生　C 末端的 ATH 衍生物在酸性条件下与 $[NCS]^-$ 反应而被裂解生成 ATH－氨基酸，新的 C 末端自动形成 TH，而不必重新活化。

因此，整个测序过程只需在开始时对 C 端进行一次性活化，并修饰 Asp 和 Glu 侧链羧基，以及 Thr 和 Ser 的羟基。实际上循环反应的只有烷基化和裂解两步（见图 9－2）。

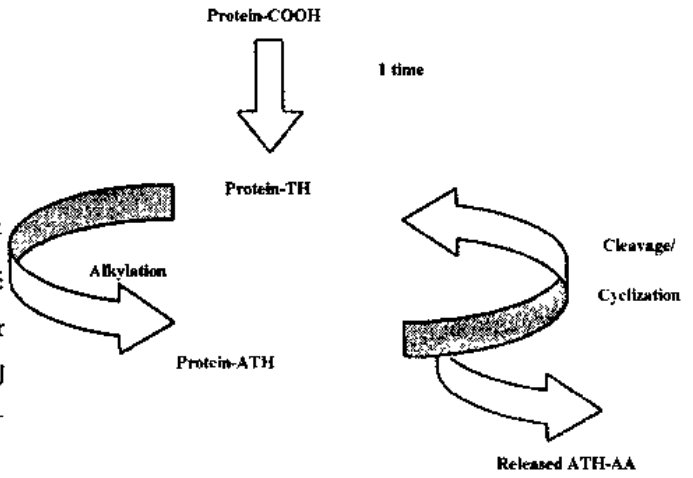


图 9－2　C 端序列分析过程示意图

（二）C 端蛋白质序列仪

近两年来，不断有商品化的 C 端序列仪问世，C 端序列仪一般由 N 端序列仪改装而成，其基本结构相似，所不同的是：①反应所用的试剂不同，因此不可兼容，否则极易堵塞管道；②在 C 端序列仪中，所用的化学反应均在弹筒型反应室中进行，切割下来的 ATH-AA 仅在转化腔内干燥和溶解后即进入 HPLC 系统分离分析。而在 N 端测序仪中，ATH-AA 需在转化腔中转化为 PTH-AA。图 9－3 是 ATH-氨基酸色谱图。

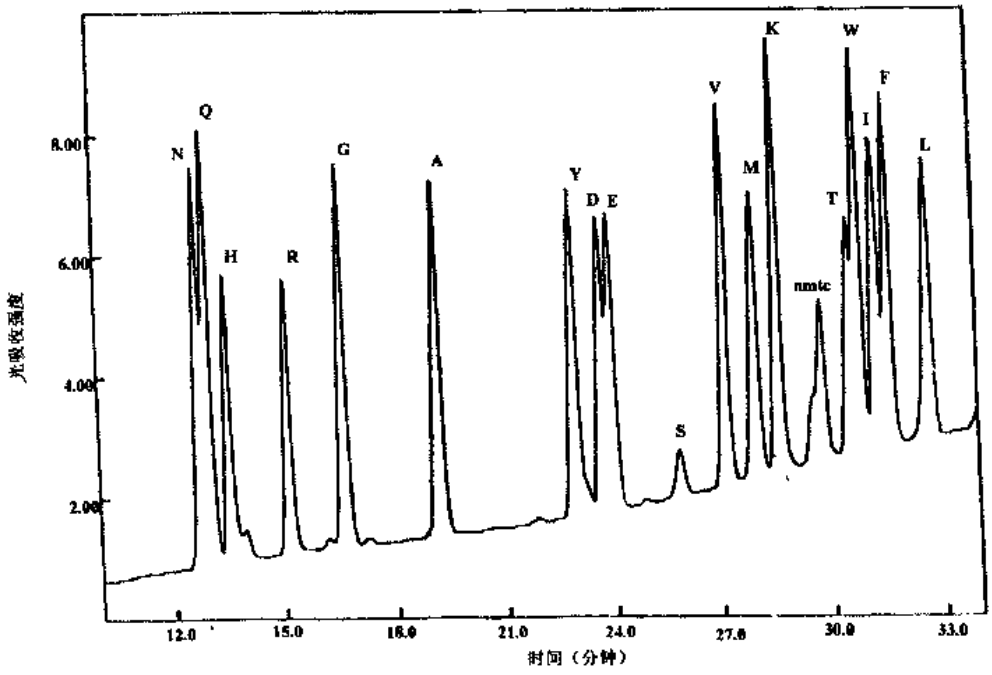


图 9－3　ATH－氨基酸色谱图

（三）C端测序样品前处理

1. 纯度鉴定　可采用SDS－PAGE、反相HPLC、毛细管电泳、阴离子或阳离子的FPLC等方法鉴定样品纯度。

2. C端脱盐的方法　目前是采用ProSorb装置，其结构和使用方法与ProSpin大致相同，所不同的是用过滤器代替了ProSpin中的离心管，其可将样品中的小分子物质随水分一起吸收，以达到脱盐和浓缩样品的目的。具体步骤为：①将插管插入套管上部，过滤器插入套管下部并接触插管底部的聚亚乙烯基二氟（PVDF）膜；②用15μL甲醇润湿PVDF膜；③在插管中加入50μL 0.1%TFA，待过滤器将其完全吸干；④将样品溶液加入，体积不少于100μL，吸干；⑤用TFA洗膜1～2次，吸干；⑥取出插管，置入恒温加热块上55～60℃加热至少10分钟。

3. 巯基修饰方法　采用丙烯酰胺和4－乙烯吡啶修饰方法。

4. ε－氨基的衍生　为了使Lys－ATH衍生物获得更好的分离效果，需先将蛋白质或多肽中Lys的ε－氨基与异氰酸苯酯（Phenyl isocyanate PhNCO）反应成衍生物。具体操作步骤为：①脱盐后，从恒温加热块上取出插管；②用3μL二异丙基乙胺（Diisoprpylethylamine DIEA）湿润PVDF膜；③在膜上加3μL PhNCO；④加热5～10分钟；⑤将②至④步重复两次。

C端测序技术仍处于研究和发展阶段，虽然该技术还不尽完善，但对于确认表达蛋白质的C端是否正确，以及判断在表达、纯化过程中是否发生了不必要的加工大有帮助。

第二节　生物质谱技术在蛋白质和多肽结构测定中的应用

已发展了一百多年的质谱（Mass Spectrum，MS）技术，由于其所具有的高灵敏度、高准确度、易于自动化等特点，已成为蛋白质和多肽结构分析的重要手段，特别是生物质谱仪的诞生使质谱技术在生命科学领域获得广泛应用和发展。

一、质谱技术的基本原理

质谱（Mass Spectrometry）是带电原于、分子或分子碎片按质荷比（或质量）的大小顺序排列的图谱。质谱仪是一类能使物质粒子电离化成离子并通过适当的电场、磁场将它们按空间位置、时间先后或者轨道稳定与否实现质荷比（即质量与电荷的比值m/e）分离，并检测强度后进行物质分析的仪器。质谱仪主要由分析系统、电学系统和真空系统组成。用于分析的样品分子（或原子）在离子源中离子化成具有不同质量的单电荷分子离子和碎片离子，这些单电荷离子在加速电场中获得相同的动能并形成一束离子，进入由电场和磁场组成的分析器，离子束中速度较慢的离子通过电场后偏转大，速度快的偏转小；在磁场中离子发生角速度矢量相反的偏转，即速度慢的离子依然偏转大，速度快的偏转小；当两个场的偏转作用彼此补偿时，它们的轨道便相交于一点。与此同时，在磁场中还能发生质量的分离，这样就使具有同一质荷比而速度不同的离子聚

焦在同一点上，不同质荷比的离子聚焦在不同的点上，其焦面接近于平面，在此处用检测系统进行检测即可得到不同质荷比的谱线，即质谱。通过质谱分析，我们可以获得分析样品的分子量、分子式、分子中同位素构成和分子结构等多方面的信息。其原理为样品分子经不同方法被电离、裂解成各种质量不同的带电正离子，在加速电场的作用下，形成离子束射入质量分析器。由于受磁场的作用，入射的离子束便改变运动的方向。当离子的速度和磁场强度不变时，离子作等速圆周运动，其轨迹与质荷比的大小有关。子是，各种离子会按其质荷比的大小分离开。然后，由记录系统记录而得质谱图。由于质谱图中谱线的位置是严格地按照与$\sqrt{m/e}$的数值成正比的规律排列的，所以，根据谱线的位置及相应离子的电荷数，就可知道其质量数，可以进行定性分析；根据谱线的黑度或相应的离子流的相对强度，可以进行定量分析。质谱仪由进样系统、离子源、质量分析器、检测器、记录系统、真空系统及计算机等部分组成（见图 9－4）。

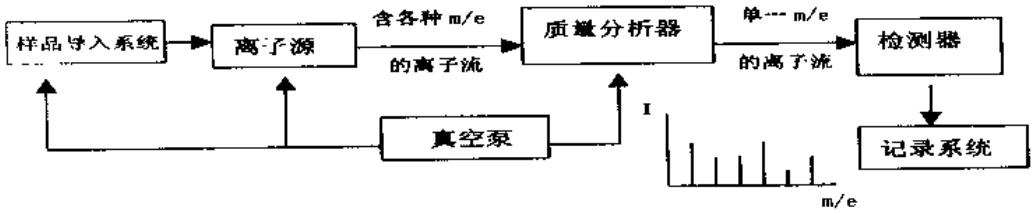


图 9－4 质谱仪示意图

二、生物质谱仪

由于生物样品极不稳定，致使电子轰击电离（EI）、化学电离（CI）等传统电离技术无能为力。为此，20 世纪 80 年代初推出了快原子轰击电离（FAB）技术，到了 80 年代末，随着两种崭新的尤其是适合蛋白质研究的软电离技术 ESI（电喷雾电离）和 MALDI（基质辅助激光解析电离）的出现，质谱成为现代蛋白质科学中最重要和不可缺少的组成部分。电喷雾电离常采用四极杆质量分析器，所构成的仪器称为电喷雾（四极杆）质谱仪（ESI－MS）；基质辅助激光解析电离常用飞行时间作为质量分析器，所构成的仪器称为基质辅助激光解析电离飞行时间质谱仪（MALDI－TOF－MS）。ESI－MS 的特点之一是可以和液相色谱、毛细管电泳等现代化的分离手段联用，从而大大扩展了其在生命科学领域的应用范围，包括药物代谢、临床和法医学的应用等；MALDI－TOF－MS 的特点是对盐和添加物的耐受能力高，且测样速度快，操作简单。此外，可用于生物大分子测定的质谱仪还有离子阱（ion trap，IT）质谱和傅里叶变换离子回旋共振（Fourier transform ion cyclotron resonance，FTICR）质谱等。而最近面市的最新型的生物质谱仪是液相色谱－电喷雾－四极杆飞行时间串联质谱（LC－ESI－MS－MS）与带有串联质谱功能的 MALDI－TOF 质谱仪，前者是在传统的电喷雾质谱仪的基础上采用飞行时间质量分析器代替四极杆质量分析器，大大提高了仪器的分辨率、灵敏度和质量范围，其商品名有 Q－TOF 和 Q－STAR 等；后者是在质谱中加入了源后降解（post－source decay，PSD）模式或碰撞诱导解离（collisionally induced

dissociation，CID）模式，从而使生物大分子的测序成为可能。

（一）快原子轰击质谱（FAB－MS）

FAB质谱也叫液体二次离子质谱（LSI－MS），它是一种用快速原子轰击被分散在高沸点溶剂（如甘油）中的待测化合物，从而产生分子离子。快原子的产生是通过电离惰性气体Ar，Xe或He产生的Ar^+，Xe^+或He^+被电场加速，从而具有较大功能，以后通过Ar气室进行电荷交换反应：Ar^+（快的）＋Ar→ Ar（快的）＋ Ar^+经电荷交换后的低动能（热）的Ar^+被偏转出快原子流，获得高动能快速的Ar原子对样品分子进行轰击，一般样品调在甘油基质之中，当快原子束轰击在涂有样品的金属板上（见图9－5），快原子大量动能以各种方式消散。

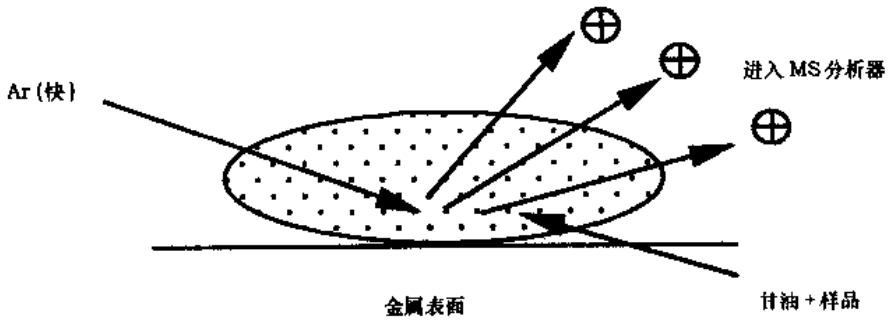


图9－5　FAB法示意图

其中的一些能量导致样品的挥发和离解。这种方法要求简单，灵敏度高。FAB产生的分子离子非常稳定，不易裂解，是准确测定多肽化合物分子量的有效方法。应用于一些较复杂的混合物，能测出其各组分的分子量。当然由于不易获得碎片峰，FAB质谱在测定多肽顺序时遇到困难。为了克服这些困难，应用了另一种叫串联质谱（MS/MS）技术。它是把从FAB出来的分子离子（MS－1）和惰性原子再一次轰击（collision－induced dissociation，CID或叫collisionally activated dissociation，CAD），从而获得一张碎片图谱（MS－2）。这样，对于一些蛋白质裂解的多肽碎片，可从MS/MS上得到分子量值及顺序信息。对于一些较复杂的多肽混合物，可先从FAB上得到每个组分的分子离子，再通过CID或CAD对其中某一个或全部离子依次分析。

（二）电喷雾质谱（ESI－MS）

电喷雾电离方法是靠强的电场使分子电离。样品溶液以很低的流速（1～20μL/min）从毛细管流出，用一个高电压（1～5kV）加在毛细管柱头上，使柱头液体雾化成很细的带电液滴，这种带电液滴在逆向干燥气流中（N_2）因挥发而使液滴表面的电荷密度增大，直接产生的库仑斥力和液滴表面张力的雷利极限值相等，此时，液滴就会爆裂，而形成一些更小的子液滴，这些小液滴又会蒸发，直至达到雷利极限再一次爆裂，这个过程一直进行下去，直到液滴变得非常小。由于液滴的曲率半径很小，它的表面电

荷密度很大，结果在液滴表面形成非常强的电场，这电场足以从液滴中解析出离子并使其进入到周围气体中。这种分子离子特点是分子离子往往带多个电荷。这些离子的形成是靠吸附上或失去若干个质子而形成，所以在正离子或负离子谱上会观察到（M + nH)$^{n+}$或（M - nH)$^{n-}$的峰。对于生物大分子，在 ESI 谱上出来的往往是一组带不同电荷的分子离子峰，根据仪器上记录下来的每个峰的质荷比及电荷数即可算出分子量值。实际上从一组峰上可算出无数个分子量值，它们互相之间会略有差别，一般需在计算机上经特定的程序处理，找出最接近实际的分子量值。电喷雾质谱的优势就是它可以方便地与多种分离技术联合使用，如液—质联用（LC—MS）是将液相色谱与质谱联合而达到检测大分子物质的目的。

（三）基质辅助激光解析电离质谱（MALDI—MS）

MALDI 技术几乎是与 ESI 同时在 20 世纪 80 年代末问世。产生的离子常用飞行时间（time—of —flight，TOF）检测器来检测，所以 MALDI 名字常与 TOF 联在一起，即叫基质辅助的激光解析飞行时间质谱仪（MALDI—TOF—MS），其电离的基本原理是将分析物分散在基质分子中并形成晶体。当用激光照射晶体时，由于基质分子经辐射所吸收的能量，导致能量蓄积并迅速产热，从而使基质晶体升华，致使基质和分析物膨胀并进入气相。MALDI 所产生的质谱图多为单电荷离子，因而质谱图中的离子与多肽和蛋白质的质量有一一对应关系。即：利用激光脉冲辐射分散在底物中的分析物使其解离形成离子，并根据不同质荷比离子在仪器无场区内飞行和到达检测器时间的不同而形成一张完整质谱图。

飞行时间质谱的基本理论如下：

$$1/2\ mv^2 = ZV$$

（m 为质量，Z 为电荷数，V 为电压，v 为飞行速度）则

$$v=\sqrt{2ZV/m} \quad t=L/v$$

（t 为飞行时间，L 为飞行管长）

$$t=L/v=L\cdot\sqrt{m/2ZV}=L\cdot\sqrt{1/2V\cdot m/Z} \qquad m/Z=2\cdot V\cdot t^2/L^2$$

（其中 V、L 为仪器常数，测得飞行时间 t 即可得到 m/Z）。该技术较关键的一步是找到合适的基质，类似于 FAB。目前已研究了几百个有机化合物作为基质，但只有几个化合物适用，这是因为基质要求与被分析物质不起化学反应，在固相中基质与被分析物形成单相，在真空中要低的蒸气压，能促进离子化，能有给出质子的化学组分。总之，MALDI—TOF—MS 选择好基质和制备好样品是分析成败的关键。要点为：制备基质溶剂一般用水，水—乙腈，或者水—乙醇混合物，其浓度为 5～10μg/μL，这依赖于基质的溶解度性质。分析物以浓度 0.1μg/μL 溶于基质的溶液中，对多肽和蛋白质样品常加 0.1%三氟醋酸水溶液，基质和被分析物的最后克分子比接近于 10000∶1 和最后体积为 0.5～2μL；样品和基质混合物被放到不锈钢 MALDI—TOF—MS 样品棒的尖端，用热空气挥发干燥或真空干燥，在干燥过程中，基质和被分析物在溶液中共沉淀。影响其实验灵敏度的因素是：一般少量样品能得到比较好 MALDI 信号，即要选择合适的基质和被分析物的克分子比，较低的杂质浓度如缓冲液中的盐等也能获得比较好的质

谱图。常用的基质是尼古丁酸或其同系物。如肉桂酸衍生物、苯甲酸衍生物和芥子酸等。

ESI 和 MALDI 质谱的主要突破点在于电离方式，它们都是采用被称为“软电离”的技术，即样品分子分离时，保留分子的完整性，不会形成碎片离子而产生稳定分子离子，因而是测定生物大分子分子量的有效方法。据目前报道，ESI 测定 75000Da 分子量的化合物精确度达 0.05%～0.01%；MALDI 测定 300000Da 分子量的蛋白质精确度达 0.1%～0.05%。这是用电泳或层析测定分子量的方法所无法比拟的。ESI 质谱是最早与 LC 连用成功的质谱仪。原则上讲，ESI 质谱与 LC 连用能检测到混合物中的所有组分，当然对于亲水性很强的小肽，如单个氨基酸或二肽，则有可能被遗漏。另外，对于一些分子量较大（>1000000Da）或疏水性强的蛋白质，MALDI 更有效。它还能有效地分析较复杂的肽混合物或物理化学性质相差较大的蛋白质混合物。只要能与所选择的底物形成稳定的分散体系的样品都能用 MALDI 进行分析。它还有个优点是不受样品所含的添加物、缓冲液或盐的影响，且灵敏度也比别的离子化方式高。当然，正如前面所提到，MALDI 在选择底物方面有一定难度。

三、生物质谱在蛋白质和多肽结构测定中的应用

质谱既可以准确测定一个未知蛋白质或者不纯蛋白质各组分的分子量，同时又可以定出较复杂的蛋白质裂解的肽片段混合物中每个肽片段的分子量及顺序，它成为蛋白质一级结构测定中的一个重要手段之一。它的优点在于快速、灵敏，所需的样品量少到只需几十个 pmol 的数量级。它能解决一些用经典的蛋白质结构测定方法难以解决的问题。如对于 N 端封闭的肽和环肽样品，不能直接用蛋白质、多肽气相顺序仪测定其顺序，而质谱法则不受其影响；再如一些很难提纯的肽或蛋白质样品，质谱可测出其各组分的结构，而经典方法的前提是必须拿到纯品。

（一）分子量测定

分子量是蛋白质、多肽最基本的物理参数之一，是蛋白质、多肽识别与鉴定中首先需要测定的参数，也是基因工程产品报批的重要数据之一。分子量正确与否往往代表着所测定的蛋白质结构正确与否或者意味着一个新蛋白质的发现。生物质谱可测定生物大分子分子量高达 40 万，精确度高达 0.1%～0.001%，远远高于目前常规应用的 SDS－PAGE 电泳与高效凝胶色谱技术等。正确测定蛋白质的分子量，可提供很多信息，如确定分子大小，从氨基酸组成分析的结果求得准确的氨基酸组成，验证已测定的一级结构是否正确等。

（二）肽谱测定

肽指纹图谱（peptide mass fingerprinting，PMF）测定是对蛋白酶解或降解后所得多肽混合物进行质谱分析的方法，对质谱分析所得肽片与多肽蛋白数据库中蛋白质的理论肽片进行比较，从而判别所测蛋白是已知还是未知。由于不同的蛋白质具有不同的氨基酸序列，因面不同蛋白质所得肽片具有指纹的特征。肽谱是基因工程重组蛋白结构确

认的重要指标，也是蛋白质组研究中大规模蛋白质识别和新蛋白质发现的重要手段。通过与特异性蛋白酶解相结合，生物质谱可测定肽质量指纹谱并给出全部肽段的准确分子量，结合蛋白质数据库检索，可实现对蛋白质的快速鉴别和高通量筛选。PMF 常和胶上原位酶解相结合，成为蛋白质组研究中必不可少的一种手段。此外根据肽段质量数变化，可对基因产品的插入、缺失、突变进行对比分析。

采用肽指纹谱的方法已对酵母、大肠杆菌、人心肌等多种蛋白质组进行了研究。对大肠杆菌经 PVDF（聚亚乙烯基二氟）膜转印的蛋白质的研究表明，三个肽片即可达到对蛋白质的正确识别。而采用原位酶解的方法对酵母蛋白质组研究的结果显示，约 90%的蛋白质被识别，其中 30 多种新蛋白质被发现，而这些蛋白质是酵母基因组研究中未能识别的开放阅读框架。研究显示，肽指纹谱的方法比氨基酸组成分析更为可靠，这是因为 MALDI 测定肽质量的准确度为 99.9%，而氨基酸组成分析的准确度仅为 90%。另外 MALDI 可以耐受少量杂质的存在，对于纯度不是很高的样品也能得到理想的结果。

（三）肽、蛋白质序列测定技术

构成蛋白质的常见氨基酸有 20 种，一段 3 个氨基酸的肽段碎片将有 8000 种可能的排列方式，4 个氨基酸将有 160000 种排列方式，即一个特定的 4 个氨基酸序列的出现概率为 1/160000。因此，即使对于一个相当大的蛋白质组来说，五、六个氨基酸残基的序列片段已具有很高的特异性。串联质谱技术可直接测定肽段的氨基酸序列，从一级质谱产生的肽段中选择母离子，进入二级质谱，经惰性气体碰撞后肽段沿肽链断裂，由所得到的各肽段质量数差值推定肽段序列，用于数据库查寻，称之为肽序列标签技术（peptide sequence tag，PST），目前广泛应用于蛋白质组研究中的大规模筛选。较之传统的 Edman 降解末端测序技术，生物质谱具有不受末端封闭的限制、灵敏度高、速度快的特点。另外，一种间接的肽序列测定技术即肽阶梯序列技术（peptide ladder sequence），通过末端酶解或化学降解，产生一组相互之间差一个氨基酸残基的多肽系列，经 MALDI-TOF-MS 鉴定后，由所得到的肽阶梯图中各肽段的分子量差值确定末端的氨基酸序列，从而用于数据库查寻。随着 FAB-MS，MS/MS，ESI-MS 和 MALDI-TOF-MS 技术的发展，用 MS 测定蛋白质顺序越来越被人们重视。一般的方法是先用 3，4 种不同的蛋白水解酶对蛋白质进行酶解，如胰蛋白酶、胰凝乳蛋白酶、嗜热菌蛋白酶和 SV8 酶等。各种酶解片段用 HPLC 等分离得到的纯肽或简单的混合肽，可用 MS/MS 测定其各组分的顺序，然后从不同蛋白水解酶酶解片段顺序，找到其重叠部分，即可得到整个蛋白质顺序。

（四）巯基和二硫键定位

二硫键在维持蛋白和多肽三级结构和正确折叠中具有重要作用，同时也是研究翻译后修饰所经常面临的问题，自由巯基在研究亚基之间及蛋白与其他物质相互作用中具有重要意义。利用碘乙酰胺、4—乙烯吡啶、2—巯基苏糖醇等试剂对蛋白质进行烷基化和还原烷基化，结合蛋白酶切、肽谱技术，利用生物质谱的准确分子量测定，可实现对二

硫键和自由巯基的快速定位与确定。这在含有多二硫键结构的活性多肽与蛋白质研究中有重要用途。

（五）蛋白质翻译后修饰

蛋白质在翻译中或翻译后由于不同功能所需会发生 20 多类修饰，其中最常见的是磷酸化和糖基化，这些修饰将影响蛋白质的电荷数、疏水性、构象和稳定性，进而影响其生物活性，因此，在基因工程重组蛋白评估时，翻译后修饰也被认为是影响活性的一个重要因素。传统的 Edman 降解技术会破坏蛋白质修饰，肼解方法虽能得到糖链并鉴定之，但却不能进行糖链的准确定位。结合肽谱和脱磷酸酶作用，目前已可以用 MALDI-TOF-MS 对双向电泳分离蛋白质磷酸化位点进行定位。凝集素可对糖链进行特异性吸附，利用这一性质，可用凝集素对糖蛋白酶解后的肽混合物进行选择性吸附，MALDI-TOF-MS 测定质量数，再用糖苷酶降解含糖肽，质谱测定后即可确定糖链的大小，结合不同的酶解方式可确定糖基化位点，选择不同的凝集素，还可确定糖链的连接方式。串联质谱技术中子离子扫描模式可以快速选择被修饰片段，然后可根据特征丢失确定修饰类型，是目前最有效的对蛋白质翻译后修饰进行识别与鉴定的分析手段。这方面的研究已有大量文献报道。鉴于修饰的多样性，目前已有人结合生物质谱技术，通过分析数千例蛋白质翻译后修饰，建立蛋白质修饰数据库 FindMod，其完善和发展必将推动蛋白质大规模鉴定的进程。

（六）生物分子相互作用及非共价复合物

蛋白质与其他生物分子相互作用在信号传导、免疫反应等生命过程中起重要作用。软电离技术的发展，促进了生物质谱在蛋白质复合物研究中的应用，目前已涉及了分子伴侣对蛋白折叠作用、蛋白/DNA 复合物、RNA/多肽复合物、蛋白质－过渡金属复合物及蛋白－SDS 加合物等多种类型的复合物的结构及结合位点的研究。

（七）蛋白质组研究

蛋白质组是指一个基因组、一个细胞或组织所表达的全部蛋白质成分。蛋白质组的研究是从整体水平上研究细胞或有机体内蛋白质的组成及其活动规律，包括细胞内所有蛋白质的分离、蛋白质表达模式的识别、蛋白质的鉴定、蛋白质翻译后修饰的分析及蛋白质组数据库的构建。质谱技术是蛋白质组研究的三大支撑技术之一。随着蛋白质组学研究的日益兴起，蛋白质的定性分析技术：如蛋白质测序、蛋白质翻译后修饰测定等相关仪器如蛋白质测序仪和 MALDI-TOF，LC/MS/MS 等已基本成熟，但是蛋白质的差异定量分析技术还基本处于空白。美国应用生物系统公司最新推出了基于 MALDI-TOF 和 LC/MS/MS 进行定量的 ICAT（Isotope Coded Affinity Tags，同位素标记亲和标签）技术，ICAT 技术可以全景式分析两种不同组织细胞：如比较正常与肿瘤，用药前与用药后组织细胞内全部各种蛋白质表达方式量的差异，并对差异蛋白进行结构鉴定。该方法的特点是：①可以结合实验设计找到相关功能的蛋白质；②易于自动化，实验样品制备可不需要 2－D 凝胶电泳分离。其机理和过程如下：分别用 D0 和 D8 试剂与同种细胞

的不同形态（如正常细胞和病变细胞）中的所有蛋白质反应（如 D0 与正常细胞，D8 与病变细胞），试剂选择性只与胱氨酸反应，然后把两种反应产物混合在一起进行酶解，用亲和色谱分离被标记肽段，进行 MALDI-TOF（Voyager DE PRO）或 LC/MS/MS（API3000）测定，如一对峰相差 8 个质量数，则为同一种蛋白质，由 D0 和 D8 峰的相对强度进行相对定量。

ESI-MS、MALDI-TOF-MS 是 20 世纪 80 年代末期发展起来的软电离技术，它们在蛋白质、多肽化学中得到了广泛的应用。国际上也刚起步，随着各国科学家的努力，这种新技术将不断完善和发展，特别是它们与 HPLC、毛细管电泳等联用，使蛋白质、多肽结构测定向更微量化、快速化发展，也许在若干年后它将取代经典的 Edman 降解测定蛋白质、多肽顺序的方法，成为蛋白质多肽顺序测定的重要手段之一，将大大促进生命科学的发展。

第三节 X 射线晶体衍射技术在蛋白质晶体结构测定中的应用

X－射线衍射法是测定蛋白质晶体结构的极其重要方法。目前还没有一种其他工具能够直接观察到蛋白质内部的原子和基团的排列。虽然电子显微镜接近于看到大分子的轮廓，但是仍然仅限于揭露分子的大小、形状、对称性和聚集状态等。通过 X－射线衍射法（X－ray diffraction method）可间接地研究蛋白质晶体的空间结构。对晶体结构的研究将帮助人们从原子的水平上了解物质。1913 年布拉格父子用 X 射线衍射法对氯化钠、氯化钾晶体进行了测定，指出分析晶体衍射图就可以确定晶体内部的原子（或分子）间的距离和排列。

一、基本原理

用 X－射线衍射法测定晶体结构是根据晶体中原子重复出现的周期性结构。当 X－射线穿过晶体的原子平面层时，只要原子层的距离 d 与入射角的 X－射线波长 λ、入射角 θ 之间的关系能满足布拉格（Bragg）方程式：

$$2d\sin\theta = n\lambda \quad (n = \pm 1, \pm 2, \pm 3, \cdots)$$

则反射波可以互相叠加而产生衍射，形成复杂的衍射图谱（见图 9－6）。不同物质的晶体形成各自独特的 X－射线衍射图。根据记录下来的衍射图谱，经过复杂的数学处理，可推知晶体中原子的分布和分子的空间结构。

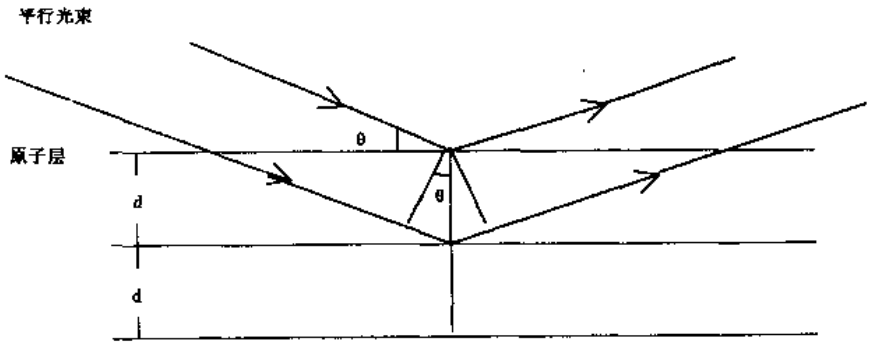


图 9－6　布拉格方程的图解

二、X 射线衍射技术在蛋白质分析中的应用

1959 年佩鲁茨和肯德鲁对血红蛋白和肌红蛋白进行结构分析，解决了三维空间结构，获 1962 年化学奖。鲍林发现了蛋白质的基本结构。1953 年克里克、沃森在 X 射线衍射资料的基础上，提出了 DNA 三维结构的模型，获 1962 年生理或医学奖。20 世纪 50 年代后期，有机化学家豪普特曼和卡尔勒建立了应用 X 射线分析的以直接法测定晶体结构的纯数学理论，在晶体研究中具有划时代的意义，特别在研究大分子生物物质如激素、抗生素、蛋白质及新型药物分子结构方面起到了重要推进作用。他们因此获 1985 年化学奖。有些蛋白质像纤维素或尼龙一样，既具有纤维状又具有晶体状，例如，丝心蛋白、角蛋白（毛发和皮肤里的蛋白质）和胶原蛋白（腱和结缔组织中的蛋白质）。德国物理学家赫佐格通过显示这些物质能够衍射 X 射线（见图 9－7），从而证明了它们的结晶度。另一位德国物理学家布里尔分析了衍射图，从而确定了多肽链中原子的间距，20 世纪 30 年代，英国生物化学家阿斯特伯里等人利用 X 射线衍射进一步了解了多肽链的结构。他们能够相当精确地计算出相邻原子之间的距离和相邻的键所成的角度。

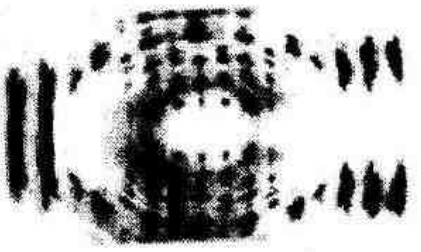


图 9－7　蛋白质的 X 射线衍射图

他们还了解到，丝心蛋白的链是完全伸展的，就是说，这些原子间键的角度能够使它们几乎排列在一条直线上。多肽链的这种完全伸展是一种最简单的排列，叫做 β 构型。但是在通常不拉伸的状态下，角蛋白分子显示出比较复杂的排列，称为 α 构型。

1951 年，加利福尼亚理工学院的泡令和科里提出，α－构型的多肽链呈螺旋形（类似螺旋楼梯）。为了弄清原子之间所有的键在未拉紧的自然取向的状态下 α－构型的排列情况，他们设计了各种模型，最后确定螺旋的每一圈具有 3.6 个氨基酸的长度，即相当于 5.4 埃。维持蛋白质 α－螺旋结构的化学力是氢键。通过化学分析和 X 射线分析确定，组成蛋白质的氨基酸都是 L－型氨基酸。

近年来，生物大分子三维结构的测定方法有很大的突破。第三代同步辐射光源的应

用使得用 20～40μm 大小的晶体解析高分辨率结构已经成为现实。多波长反常散射方法较好地解决了晶体衍射中的相位问题，一些大而复杂的蛋白质，DNA，RNA 及其复合物，特别是一些病毒，亚细胞器精细结构得到解析。此外，用晶体电子衍射方法解析膜蛋白质结构的分辨率已达 3 埃。蛋白质三维结构测定出现前所未有的高速增长的态势。目前存入蛋白质数据库的生物大分子结构模型的数目呈抛物线增长。截至 2001 年 7 月，PDB 数据库中的三维数已达 1.5 万个，其中蛋白质的比例为 89.4%。X 射线晶体衍射法测定的结构数目占 82.6%。

用 X 射线衍射晶体结构分析法研究重要生物活性大分子的三维结构、分子识别和作用机理，属于新兴的化学、生物和物理的交叉学科。结构生物学领域，是当代生命科学领域内一个十分重要的分支，而 X 射线晶体结构分析是结构生物学最主要的研究手段。

进行 X 射线衍射谱的分析，通常高纯度的材料是必须的，首先实验样品应该仔细结晶，去获得高品质的适合 X 射线衍射研究的晶体。小分子体系的结晶通常可用一般化学手段获得，然而对大分子体系如蛋白质则要求特殊技巧来产生结晶。即使可以得到合适的结晶，大分子体系结构的测定仍然比小分子要困难得多。这是因为数目众多的大分子体系结构中的原子结构确定很困难，同时对仪器要求的复杂性及数据分析都需要消耗大量的计算时间。众所周知，X 射线衍射必须在分子固相中进行，由此获得的结构信息可能就会与分子体系在生物活性状态的情况有所不同。近年来，科学家们努力地发展建立了结晶学软件，大大加快了蛋白质结构测定的步伐。以前需要几周甚至几个月才能完成的工作，现在有可能在几小时到几日内完成。

第四节 核磁共振波谱技术在蛋白质结构测定中的应用

X 射线晶体衍射在蛋白质研究中受到一定限制，主要有两方面原因，一是因为要求用于结晶的样品必须具有相当高的纯度，二是因为从晶体得到的三维结构信息可能与分子体系在生物条件下的溶液中结构信息不一致，因而确定蛋白质分子在溶液中结构信息广为使用的方法是核磁共振波谱技术，它使用较低强度辐射能测定蛋白质分子在流动化液态下的三维结构信息，从而更真实地反映蛋白质在生物环境下的结构信息，并能确定小分子与大分子复合物的结构。

一、NMR 的基本原理

(一) 原子核的一般特性

核磁共振是研究处在静磁场中的原子核与电磁波相互作用的学科，因此首先简单介绍原子核的一般特性。各种不同的原子具有不同的原子核。每种原子核的最基本的特性是它的电荷和质量。核的电荷决定于核中质子的数目；而核的质量则决定于核中质子数目与中子数目之和；而且还决定子核中粒子之间的结合能。

实验证明，根据原子核的电荷和质量这两个特性，一般情况下就可以解释原子核与

周围粒子的强相互作用，例如裂变、聚变等。但是仅仅根据这两个性质还不能解释一些弱相互作用，例如，原子光谱的精细结构、核磁共振现象等。因此还必须引入一些原子核的其他特性。在核磁共振的研究中常用的原子核特性是**原子核的自旋角动量和磁矩，原子核的电四极矩**也很重要。下面分别介绍这三个参数。

1. 原子核的自旋运动　原子核具有一定的质量和一定的体积，可以把它看成是一个接近球形的固体。实验表明，大多数的原子核如同陀螺一样，都围绕着某个轴作自旋运动。例如，常见的 H_1^1 和 C_6^{13}（6 是质子数即原子序数，也是电荷数；13 是质量数＝质子数＋中子数）核等都具有这种运动。原子核的自身旋转运动称为核的自旋运动。一般用“自旋矩”、“自旋动量矩”或“自旋角动量”来描述这种自旋运动（见图 9－8）。

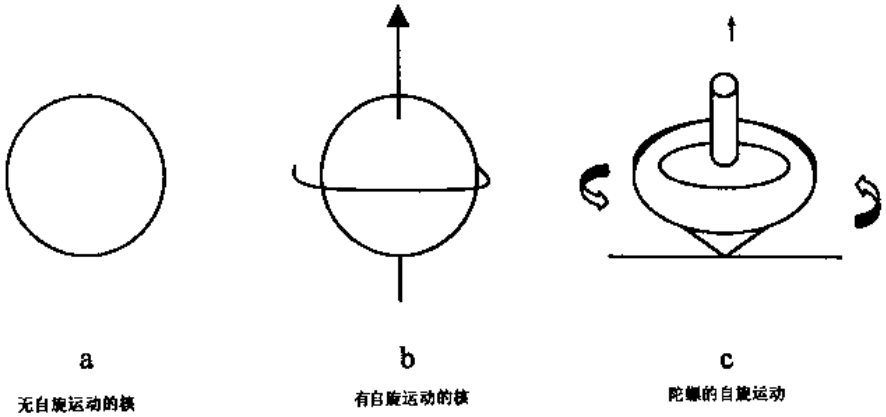


图 9－8　原子核的自旋运动示意图

根据力学中的定义，相对于某个点或某个轴的作用称为矩。它等于某点到作用点的矢量（具有一定数值和方向的量）半径 r 与作用矢量的矢量积（即叉乘积）。如：相对于某个点的作用力称为力矩，它等于 $r\times F$；而相对于某个点的运动称为动量矩，它等于 $r\times mv$。上两式中 F 表示作用力；m 表示运动物质的质量；v 表示运动的速度。mv 是描写这个运动的量，所以称为动量。当物体围绕某个轴做旋转运动时也具有一定的动量矩，常常把这种动量矩称为角动量。当物体自身旋转时产生的动量矩，称为自旋动量矩，或称为自旋角动量，也称为自旋矩。

进行自旋运动的原子核都具有一定的自旋角动量。核的自旋角动量是个矢量，用 P 表示，它的方向与旋转轴相重合。根据量子力学计算，原子核自旋角动量的绝对值由下式决定：

$$|P|=\frac{h}{2\pi}\sqrt{I\ (I+1)}=h'\sqrt{I\ (I+1)}$$

式中h—普朗克常数＝6.624×10^{-27}尔格·秒

$h'=h/2\pi$　　I—核自旋量子数，$I=0$，$1/2$，1，$3/2$，…

由上式可见，核自旋角动量的大小决定于核自旋量子数 I。它表示某种原子核所固有的特性，不同的原子核有不同的 I 值。它只能取零，半整数，整数。是不连续的，故为量子数。实验证明：①$I=0$：原于核的质子数与中子数都为偶数，则它的质量数也为偶数。这种核的自旋量子数 I 等于零。即核的自旋角动量为零，不产生核磁信号。如：C_6^{12}，O_8^{16}，S_{16}^{32}，Se_{34}^{80}；②$I=1$：原子核的质子数与中子数都为奇数，它的质量数为偶数，则 I 是整数，这类核有自旋现象，如：H_1^2，N_7^{14} 等；③$I=1/2$：原子核的质子数和中子数其中有一个是奇数，它的质量数为奇数，则 I 是半整数（1/2，3/2，5/2，…），这类核是核磁共振研究的主要对象，如：H_1^1，C_6^{13}，F_9^{19}，N_7^{15}，P_{15}^{31}等。

2. 原子核的磁矩　原子核带有一定的正电荷，粗略地讲，可以认为这些电荷均匀地分布在原于核的表面上。$I=0$ 的核没有自旋运动。$I\neq0$ 的原子核有自旋运动，故这些电荷也围绕着旋转轴旋转，从而产生一个循环电流。有循环电流就会产生一个磁场，这就如同电流流过线圈时产生磁场一样。即凡是自旋不为零的原子核都会产生一个磁场，也就是说它们像一个小磁铁一样具有磁性质。一般是用磁矩来描述这种性质。核磁矩 μ 的大小取决于核的自旋量子数 I 及磁旋比 γ，即　$\mu=I\cdot\gamma\cdot h/2\pi$

3. 原子核的电四极矩　前面曾提到，有一些原子核中的电荷分布是球形对称的，但大多数原子核中的电荷分布不是球对称的。一般是用原子核的电四极矩来度量核中电荷分布离开球形对称的程度。核中电荷分布的情况接近于旋转椭球体的形状。用 Q_N 表示核的电四极矩，它的公式为：$Q_N=2/5Z_e(b^2-a^2)$　式中 b 和 a 为椭球体的两个半径；Z_e为核的电荷。

（二）核磁共振条件

1. 量子力学观点　前面讲到，自旋角动量不等于零的原子核都具有自旋磁矩。如果把这个核放到静磁场中，那么磁场对核磁矩有一个作用力，使磁矩在磁场中具有一定的能量。核磁矩在静磁场中的能量决定于核磁矩在静磁场方向上的投影与磁场强度乘积的负值。

$$E=-\mu_z\mathbf{H}_0 \quad 又因为\ \mu_z=g_N\mu_N M \quad (M=I,\ I-1,\ \cdots,\ -I)$$

g——兰德因子

M——电子的轨道磁量子数

即核磁子在 z 轴上的投影是一些不连续的数值，这表示核磁矩 μ 在空间的方向不是连续变化的，而只能取某几个固定的方向。这就叫核磁矩的空间量子化。将两式合并得：

$$E=-g_N\mu_N MH_0$$

可见核磁矩在磁场中的能量是量子化的，我们把这些不连续的能量值称为原子核的能级。按着能量值的大小画出的图，称为能级图。核磁矩在磁场中的能级数目决定于核自旋量子数 I，能级总数为（2I+1）。

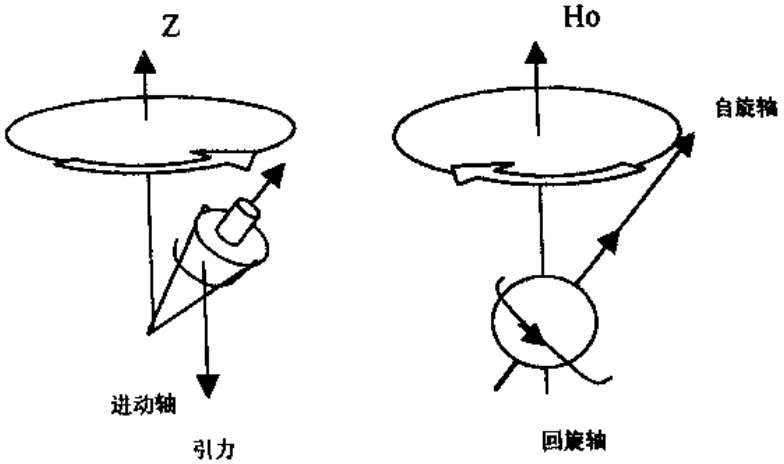
根据量子力学理论，如果我们把电磁波作用到原子核系统上，当这个电磁波的频率所决定的能量的量子 hυ，正好等于原子核两个相邻的磁能级之间的能量差额（$E_{M\pm1}-E_M$）时这个电磁波就会引起原子核在这两个能级之间的跃迁，从而引起原子核对电磁波的吸收或者辐射。我们可以通过观察原子核对电磁波的吸收现象来观察这种共振跃迁。这就是核磁共振现象。得核磁共振条件是：

$$h\nu = g_N \mu_N H_0$$

h—普朗克常数；ν—电磁波的频率

由公式可见，共振频率与磁场强度成线性关系。可以固定磁场强度改变频率，或固定频率改变磁场，以满足共振条件。

2. 经典力学观点　前面讲到，原子核如同陀螺一样作自旋运动，因而具有一定的自旋角动量。如果陀螺不旋转，当它的轴线离开垂直于地面的方向时，在重力的作用下，它就会倒下来。如果陀螺旋转起来，当它的旋转轴离开垂直于地面的方向时，在重力的作用下，它并不倒下来，而是使它的旋转轴围绕着地面垂直方向进行转动，从而描绘出一个圆锥形轨迹。这就是陀螺在地球引力场中的进动。与此类似，具有自旋角动量的核同时也具有磁矩。如果把核放在静磁场当中，磁场对磁矩有一个作用力，这就如同指南针在地球磁场的作用下永远指向南北方向一样。但是与指南针运动（无自旋运动）不同的是，由于原子核具有自旋运动，磁场对于核磁矩的作用力不是使核磁矩向磁场方向运动，而是使核磁矩围绕磁场方向进行转动，从而描绘出一个圆锥形轨迹。这就是原子核在静磁场中的进动（见图 9－9）。



a 陀螺在引力场中的进动　　b 核在静磁场 Ho 中的进动

图 9－9　原子核在静磁场中的进动

原子核进动的速度用角速度或角频率来描述，它与磁场强度成正比。其关系可用 Larmor 方程说明：$\omega_0 = \gamma/2\pi \cdot H_0$　γ—磁旋比（核磁矩与核角动量的比值）或 $\omega_0 =$

γH_0当发生核磁共振时，照射无线电频率必须等于原子核的进动频率 $\nu=\nu_0$ $\omega=\omega_0$在外磁场中若使核发生自旋能级跃迁，所吸收的光子能量 $h\nu=\Delta E$，于是

$$\nu=\gamma/2\pi \cdot H_0 \qquad \omega=\gamma H_0$$

即为核磁共振条件，亦为核磁共振基本公式。

上式表明，共振频率与磁场强度成线性关系，这一结论与量子力学得出的结论一致。

（三）NMR 技术常用的参数

1. 化学位移（chemical shift） 发生核磁共振时，如果有机化合物的所有质子的共振频率都相同，核磁共振谱上就只有一个吸收峰，这样，核磁共振技术对化学家就毫无意义。但事实并非如此，在 1946 年发现对于溶液或液体 N^{14} 的吸收频率取决于核的化学环境。这一特征很快被证明是一个普遍的规律。我们知道任何原子核都不是裸核，它的外围被电子云包围，不同类型的核，核外围电子云密度不同。换句话说，每一个化合物都含有不同的基团，以质子为例，不同类型的氢核由于在化合物分子中所处的化学环境不同，受到的实际磁场也不同，因此，每一种化学环境不同的质子都可观察到不同的信号。其实质是核的化学环境不同，拉摩尔进动频率不同。我们把由于原子和分子中核的化学环境不等价而引起拉摩尔进动频率的变化叫做化学位移。用 δ 表示，单位为 ppm。它的存在使核磁共振波谱成为测定分子结构的有用工具。对化学位移的测试可以了解价电子对核的屏蔽作用，从而鉴定分子中各种化学基团的存在。其表示公式为：

$$\delta=\frac{\upsilon_{样品}-\upsilon_{标准}}{\upsilon_{标准}}\times 10^6 \ (\text{ppm})$$

2. 自旋—自旋偶合（spin－spin coupling） 乙醇在仪器分辨率不太高的情况下，可侧得三个单峰。由此似乎可以认为化学上等价的原子核都同时发生共振，因而对应着一个单峰。但实际上，如果用分辨率稍高的仪器记谱时发现，有些单峰变成了多重峰。以乙醇为例，CH_2 分裂为四重峰；CH_3 分裂为三重峰。谱线的这种精细结构不是由于原子核化学环境的不等价而引起的，即它不是化学位移。产生这种精细结构的原因是由于分子内部原子核的磁相互作用引起的。但这种作用不是直接发生的，而是通过电子壳层传递的。分子中的核首先在电子壳层中感应出一个磁矩，然后再作用到被侧核上，从而改变了被侧核的共振条件，使谱线分裂为多重线。因此这种作用称为自旋—自旋偶合，谱线的精细结构称为自旋—自旋分裂。

自旋—自旋偶合与化学位移有很大的不同。它的特点是：①各个谱线分量之间的距离相等；②谱线分量的间距与外磁场强度无关；③谱线各个分量的强度是对称分布的；④如果对它发生作用的原子核（$I=1/2$）数目为 n，那么谱线就分裂为（$n+1$）条多重线。对于 $I>1/2$ 的核，谱线分裂数目为（$2nI+1$）。以上乙醇的例子是同一种原子核之

间的相互作用，称为同核偶合；如果在不同的原子核之间发生相互作用就称为异核偶合。

自旋—自旋偶合指标可以用来研究分子的超精细结构，以及各种基团的相互作用、作用力和空间结构。

3. 弛豫时间（relaxation time）　核自旋系统由于受到外界作用离开平衡状态以后，能够自动地向平衡状态恢复，即从高能级恢复到基态的过程称为弛豫过程。弛豫现象是由于物质相互作用而发生的，弛豫时间是描述这种相互作用的参数。T_1 称为纵向弛豫时间，是描述核磁化强度 M 的纵向分量的恢复过程及自旋与晶格（周围介质）的相互作用，因此也称为自旋—晶格弛豫时间；而 $T2$ 称为横向弛豫时间，描述 M 的横向分量恢复的过程及自旋与自旋相互作用，因此也称为自旋—自旋弛豫时间。所以准确地测量弛豫时间对研究物质相互作用的机理，研究物质结构与性质都有很重要的意义并可作为测定生物活性物质的运动及能量转换特征等有用指标。

（四）核磁共振波谱仪

从 20 世纪 50 年代中期到现在已经研制出了种类繁多的 NMR 商品仪器。它们可分为两大类：常规谱仪和做研究工作的谱仪。按其工作方式又可分成连续波谱仪（CW－NMR）和脉冲傅立叶变换谱仪（PFT－NMR），后者比前者设计更复杂，功能更强。它们所具有的共同部分是：磁铁、磁场稳定单元、探头、射频发射与接收单元、波谱显示与记录单元等。磁铁是 NMR 谱仪的心脏，其要求磁铁能产生尽量强的、非常均匀、非常稳定的磁场。目前采用的有三类磁铁：永久磁铁、电磁铁和超导磁铁。永磁的特点是，不需要磁铁电源和冷却系统，运转费用低，而且具有优异的长期稳定度，但因磁场固定，不能在宽范围内改变，并受温度影响较大，一般用于常规仪器。电磁场为了产生所需磁场它需要很稳定的电源和磁铁冷却系统且要保持磁铁温度恒定，它可以通过改变磁铁电流在较广范围内改变磁场。超导磁铁的磁场强度可大幅度提高，现已生产出 900MHz 的谱仪。因超导线圈处在液氦之中，温度很稳定且处在超导状态时，其电阻等于零，故在线圈中可以维持一个很稳定的大电流，从而产生一个稳定的磁场。NMR 仪的简单原理如图 9－10.

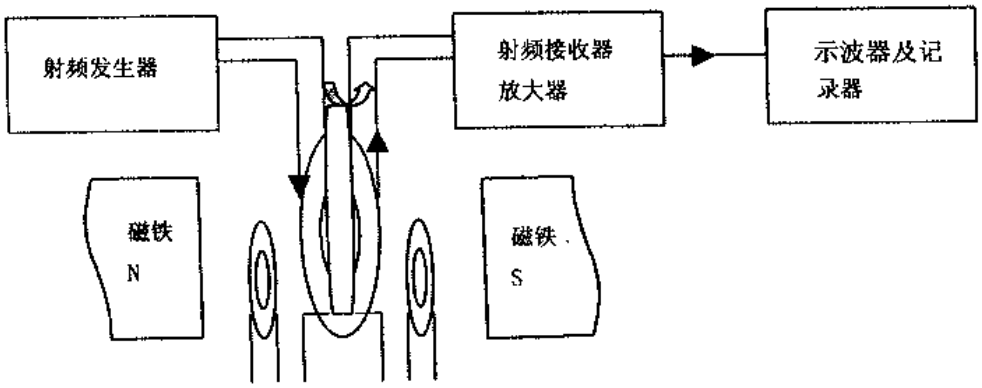


图 9－10　核磁共振仪示意图

图 9－10 中的两个磁铁产生均匀、稳定的静磁场 H_0，两磁极间放样品管，并有一机械旋转装置，可使样品管在两极间旋转。样品管外面是射频振荡线圈，它被安装在与外磁场 H_0 垂直的方向，可发射固定频率的无线电波如：60，80，90，100，200MHz 等。该磁场的有效部分是一个强度为 H_1 绕着 H_0 旋转的磁场，当射频频率等于核的进动频率（$\omega=\gamma H_0$）时，发生共振。把得到的信号送到射频接收器，经射频放大器放大，中频检波放大，音频相检波放大，通过相位调节，得到所需要的纯吸收型或色散型信号，并由记录器给出谱图。y 轴为共振峰强度，x 轴为磁场强度。共振信号强度与在该特定强度磁场下共振的核数成比例，因此，可由峰面积直接测量核的数目，目前大多数仪器备有自动积分仪。

生物 NMR 研究需要采用超导高分辨率的谱仪，场强尽可能高。一般采用 500MHz 以上的谱仪。谱仪射频发生器和去偶器要求频率范围宽，射频脉冲的频率、相位和功率的调节要精细。不同射频通道产生的脉冲在相位上要相干。功率放大器要求功率大而稳定，并可程控地衰减和传输。

在进行生物 NMR 实际操作时，一般的步骤为：样品的温度控制、样品的锁场、探头调谐、匀场调节、脉冲锁定等。

在进行实验数据参数设定时，主要考虑以下几点：中心频率设定、谱宽设定、数据点设定、接收器增益设定、水峰抑制等。影响谱峰质量的环境因素：地基震动造成磁场调制、室温的变化、磁场漂移等这些因素都将使噪音增大。

二、生物 NMR 常用的实验技术

近代生物 NMR 研究一般采用二维 NMR 实验技术，对于更复杂的生物大分子，还可采用更先进的三维和四维异核 NMR 实验技术，它需要同位素 N^{15} 和 C^{13} 标记的样品，目前国内还不能普及，因此这里介绍用于生物大分子溶液三维结构测定的同核二维 NMR 实验技术。

（一）同核相关谱

同核相关谱（COSY）用来鉴定具有偶合关系的质子对。其形态为矩形图谱中间分布着一些同心圆，实际上是一些吸收峰的等高线截面图，称之为交叉或相关峰，交叉峰的出现表示相应质子间存在着偶合关系。二维与一维谱相比，结果简单明了，分辨率高，信息量大，在一张谱图就可以显示全部质子间偶合关系。其缺点是：交叉峰与对角峰的相位差 90°，如果交叉峰为吸收线形，对角就为色散线形，因此普通 COSY 谱只能以绝对值方式表示，谱分辨率受影响，且缺少精细结构信息，灵敏度较差，故不多用。

（二）双量子滤波相关谱

双量子滤波相关谱（DQF－COSY）就是为了克服以上缺点而建立的。所谓滤波就是采用特殊的相循环，检出特定量子阶跃的磁化迁移而除去其他跃迁。它用于检出两个或两个以上自旋间的偶合相关，滤去不参与偶合的信号（主要是对角峰）。因此能明显改善靠近对角线的交叉峰的识别。DQF－COSY 谱一般采用所谓时间相位比例增量法

（TPPI）记录谱图，即所谓的相敏 COSY 谱。其每一个交叉峰都由若干个正负相间的信号组成，其所构成的精细结构包含有质子间复杂的偶合关系和偶合常数等进一步的信息。

（三）全相关谱

全相关谱（TOCSY）　又称 HOHAHA 谱，在它的脉冲序列中包含有自旋锁定的步骤，在有自旋锁定的条件下，相偶合的自旋之间会产生交叉极化转移的现象。对于有不止两个自旋的体系，取决于各向同性混合时间的长度，磁化矢量可沿偶合链传递，如果各向同性混合时间足够长，磁化矢量可从一端传到另一端，因此能将一个自旋体系的全部自旋相关联。TOCSY 谱的优点是相关峰所有分量都具有相同的相位，不会产生分量之间的自抵销。并且交叉峰和对角峰都具有吸收型线形。因此，TOCSY 谱具有比 COSY 谱更高的灵敏度。现广泛用于自旋体系（残基类型）的识别。

（四）NOE 相关谱（NOESY）

NOE 相关谱（NOESY）用于鉴定质子在空间的相互接近。空间上相互接近的质子之间，会产生交叉弛豫的现象，即 NOE 效应。其谱的形态与 COSY 谱十分相近，但是谱中的交叉峰不是表示质子间通过化学键的相关，而是质子间在空间的相互接近。NOE 能提供质子间距离的信息，质子间距离可作为几何约束用于生物大分子结构计算，因此是 NMR 技术测定生物大分子三维结构的重要参数。由于化学交换也会产生 NOE 效应，NOESY 实验用的脉冲序列无法区别交叉弛豫和化学交换产生的磁化迁移。在生物大分子中，这两种机制产生符号与对角峰相同的交叉峰的区别还得采用 ROESY 实验。

（五）ROESY 谱

这是在一种旋转坐标下测定的 NOESY 谱。采用类似于 TOCSY 实验中自旋锁定的脉冲序列，它对于上述两种机制产生不同的结果，交叉弛豫产生的 NOE 符号与对角峰相反，而化学交换产生与对角峰符号相同的交叉峰。

三、NMR 在蛋白质研究中的应用

NMR 在生物学中的应用具有下述优越性：①被测的生物样品不被破坏；②能测定与生物状态接近的溶液中生物高分子的结构与浓度；③研究生物样品的动力学过程及其相应的能量变化规律；④能获得多种参数，为分析生物分子的结构提供信息；⑤可测定多种原子核，如，H^1，C^{13}，P^{31}，N^{14}等。

（一）研究生物大分子的结构

分子生物学的核心课题之一，就是研究蛋白质和核酸等生物大分子的结构与功能。蛋白质和核酸等生物大分子都有一级、二级和高级结构。以往分析大分子的高级结构主要是靠 X 射线晶体衍射仪来进行结构分析，但结晶过程会改变分子的三级结构。利用

一维 NMR 技术分析生物分子的结构也主要局限在中等分子量大小的一级和二级结构。但在 Ernst 发展了二维核磁共振即 2D—NMR 以后，NMR 就能测定溶液中生物大分子的高级结构。NMR 是目前测定溶液中生物大分子高级结构的唯一技术。

（二）蛋白质折叠和结构转换的研究

蛋白质从具有一级结构的肽链如何形成具有生物功能的三级结构？这一问题是 DNA 到生物功能的蛋白质之间唯一还没有解决的问题，具有重大的理论意义。蛋白质的溶液构象，折叠途径（中间态理论），为揭示新生态肽链的折叠过程打下基础。蛋白质的较低级的结构是形成更高级结构的基础；而高级结构可能对低级结构有进一步的调整和完善作用。其中，二级结构和折叠中间体的形成直接影响着球状蛋白质的折叠和寡聚体的组装，结构转换和错误折叠往往导致蛋白质的聚合。二级结构转换已引起人们的重视，特别是某些蛋白质中的 α—螺旋向 β—折叠转变是改变蛋白质稳定性和聚合的原因，寡聚体组装和蛋白质的聚合遵循不同的分子机理。前者有一定的专一性，而后者则是蛋白质分子的随机集合。因此，选择寡聚体蛋白质折叠和装配这一复杂生物学问题进行研究，将能较深入地揭示蛋白质折叠和聚合问题。这一方面的研究将用园二色性、荧光、微分差热（DSC）谱以及 NMR（特别是采用 N^{15}，C^{13} 和 F^{19} 标记蛋白质的异核多维 NMR 等）方法研究中等分子量大小有重要生理功能的蛋白质的溶液构象和结构转换。折叠/去折叠方面将逐步由体外系统转向体内系统。

（三）蛋白质与蛋白质之间的相互作用的研究

蛋白质折叠的研究主要是研究分子内的相互作用，面蛋白质和蛋白质之间的相互作用是研究分子间的相互作用。它们之间有内在的共性，但是由于研究的对象不同，也有各自的特点。蛋白质和蛋白质之间的相互作用包括亚基的相互作用，抗原和抗体的相互作用，一些蛋白质和受体的相互作用等等，揭示它们的本质规律有重要的理论意义，同时，它也是分子设计和药物开发的结构基础。利用生物 NMR 技术着重在不同类型亚基的相互作用。具体内容包括建立蛋白质—蛋白质相互作用的数据库，分析结合区的结构特征，构象模块的确认，自由能的分析，热力学和动力学的研究和比较蛋白质分子内识别与分子间识别的差异性。

（四）蛋白质三维结构和构效关系的研究

采用近代 NMR 波谱技术，系统研究同源蛋白质的溶液三维结构，并在此基础上探讨其构效关系，进一步揭示了这类毒素与不同靶受体结合特异性和选择性的分子基础，完成了两个钠离子通道抑制剂和两个钾离子通道抑制剂的溶液三维结构测定。

（五）生物体系内代谢过程的研究

NMR 可以用来测定完好的生物组织中分子的浓度。结合同位素标记利用 C^{13}，N^{15}，H^{2} 等波谱来推断生物体系内的代谢过程，研究与代谢有关的分子病。例如，血红蛋白是存在于高等动物红细胞中的一种特殊的蛋白质，起着输送氧气的功能，血红蛋白

的结构与功能的演变，生物的遗传、进化与某些分子病有关。利用 NMR 技术研究水溶液中血红蛋白的波谱时，发现血红蛋白的色氨酸残基上，可交换的—NH 共振峰在氧化态和还原态之间是不同的。从氧化态到还原态这些谱线的变化规律，可以用于探讨血红蛋白在不同机能状态下的结构特征。对血红蛋白 A 等的 α 链和 β 链波谱进一步研究，发现广泛的结构变化仅发生在一个链的某部位上，与邻接血红素基团的链上只有微小的变化。波谱研究表明，血红蛋白的不规则性与导致它的机能失常有关。

（六）蛋白质空间特定区域的动力学过程研究

NMR 是研究酶的催化反应和配体与生物大分子结合的动力学非常有效的技术手段之一。可以从化学交换和浓度随时间的变化得到有关动力学的信息。交换过程的研究，首先要确定交换过程是一级还是二级交换。有效的方法是 2D 磁化转移实验。如，用 2D 交换谱研究了二氢叶酸还原酶—叶酸—$NADP^+$ 复合物与游离的 $NADP^+$ 和叶酸全部交换过程。有关小分子与蛋白质结合的电荷状态和内部复合物构型的动力学，不能从 X 一射线单晶衍射获得，但却很容易从 NMR 获得。分子的离子状态可以清楚而且简单地通过适当的核的化学位移指出，因为化学位移对分子的电荷状态是敏感的。在蛋白质—配体复合物中，当蛋白质链中含有组氨酸和苯丙氨酸时，复合物的芳香环常常翻转，这些翻转可以用来提供蛋白质空间特定区域的动力学性质。

生物物理学、生物化学与分子生物学、遗传学、细胞学、生理学甚至分类学以及医药学等都已应用 NMR 技术去进行各种有意义的探索，并力图推动这些学科的研究工作向分子水平发展。

（刘师莲）

第十章　生物信息学

第一节　概　　述

随着基因组学和蛋白质组学研究的进一步深入，生物信息学的理论和技术越来越受到人们的关注并得到了进一步的发展。生物信息学是综合运用生物学、数学、物理学、信息科学以及计算机科学等诸多学科的理论方法的一种交叉学科。

从目前生物信息学的研究情况来看，国际上公认的生物信息学的研究内容，大致包括以下几个方面：①生物信息的收集、存储、管理与提供：建立国际基本生物信息库和生物信息传输的国际联网系统；建立生物信息数据质量的评估与检测系统；生物信息的在线服务；生物信息可视化和专家系统。②基因组序列信息的提取和分析：包括基因的发现与鉴定；基因组中非编码区的信息结构分析，提出理论模型，阐明该区域的重要生物学功能；进行模式生物完整基因组的信息结构分析和比较研究；利用生物信息研究遗传密码起源、基因组结构的演化、基因组空间结构与 DNA 折叠的关系以及基因组信息与生物进化关系等生物学的重大问题。③功能基因组相关信息分析：包括与基因组信息相关的核酸、蛋白质空间结构的预测和模拟，以及蛋白质功能预测的研究等。④生物大分子结构模拟和药物设计：包括 RNA 的结构模拟和反义 RNA 的分子设计；蛋白质空间结构模拟和分子设计；具有不同功能域的复合蛋白质以及连接肽的设计；生物活性分子的电子结构计算和设计；纳米生物材料的模拟与设计；基于酶和功能蛋白质结构、细胞表面受体结构的药物设计；基于 DNA 结构的药物设计等。⑤生物信息分析的技术与方法研究：包括发展有效的能支持大尺度作图与测序需要的软件、数据库以及若干数据库工具，诸如电子网络等远程通讯工具；改进现有的理论分析方法；创建一切适用于基因组信息分析的新方法、新技术，用于解释基因组的信息，探索 DNA 序列及其空间结构信息的新表征；发展研究基因组完整信息结构和信息网络的研究方法等；发展生物大分子空间结构模拟、电子结构模拟和药物设计的新方法与新技术。⑥应用与发展研究：汇集与疾病相关的人类基因信息，发展患者样品序列信息检测技术和基于序列信息选择表达载体、引物的技术，建立与动植物良种繁育相关的数据库以及与大分子设计和药物设计相关的数据库。利用生物信息学方法进行结构功能预测要注意的是同一问题采用不同算法，可能产生相同或不同的结果。本章针对与目前分子生物学实验数据分析密切相

关的生物信息学策略及实用工具进行扼要介绍，文中涉及问题的更详细信息可参考相关网站。

第二节　生物信息数据库

近年来大量生物学实验的数据积累，形成了当前数以百计的生物信息数据库。它们各自按一定的目标收集和整理生物学实验数据，并提供相关的数据查询、数据处理的服务。随着因特网的普及，这些数据库大多可以通过网络来访问，或者通过网络下载。一般而言，这些生物信息数据库可以分为一级数据库和二级数据库。一级数据库的数据都直接来源于实验获得的原始数据，只经过简单的归类整理和注释；二级数据库是在一级数据库、实验数据和理论分析的基础上针对特定目标衍生而来，是对生物学知识和信息的进一步整理。国际上著名的一级核酸数据库有 Genbank 数据库、EMBL 核酸库和 DDBJ 库等；蛋白质序列数据库有 SWISS－PROT，PIR 等；蛋白质结构库有 PDB 等。国际上二级生物学数据库非常多，它们因针对不同的研究内容和需要而各具特色，如人类基因组图谱库 GDB、转录因子和结合位点库 TRANSFAC、蛋白质结构家族分类库 SCOP 等等。下面将简要介绍一些著名的有特色的生物信息数据库及其应用。

一、基因和基因组数据库

（一）Genbank

Genbank 库包含了所有已知的核酸序列和蛋白质序列，以及与它们相关的文献著作和生物学注释。它是由美国国立生物技术信息中心（NCBI）建立和维护的。它的数据直接来源于测序工作者提交的序列；由测序中心提交的大量 EST 序列和其他测序数据；以及与其他数据机构协作交换数据而来。Genbank 每天都会与欧洲分子生物学实验室（EMBL）的数据库，和日本的 DNA 数据库（DDBJ）交换数据，使这三个数据库的数据同步。Genbank 的数据可以从 NCBI 的 FTP 服务器上免费下载完整的库，或下载积累的新数据。NCBI 还提供广泛的数据查询、序列相似性搜索以及其他分析服务，用户可以从 NCBI 的主页上找到这些服务。Genbank 库里的数据按来源于约 55 000 个物种，其中 56％是人类的基因组序列（所有序列中的 34％是人类的 EST 序列）。每条 Genbank 数据记录包含了对序列的简要描述，它的科学命名，物种分类名称，参考文献，序列特征表，以及序列本身。序列特征表里包含对序列生物学特征注释，如：编码区、转录单元、重复区域、突变位点或修饰位点等。所有数据记录被划分在若干个文件里，如细菌类、病毒类、灵长类、啮齿类，以及 EST 数据、基因组测序数据、大规模基因组序列数据等 16 类，其中 EST 数据等又被各自分成若干个文件。

1. Genbank 数据检索　NCBI 的数据库检索查询系统是 Entrez 。Entrez 是基于 Web 界面的综合生物信息数据库检索系统。利用 Entrez 系统，用户不仅可以方便地检索 Genbank 的核酸数据，还可以检索来自 Genbank 和其他数据库的蛋白质序列数据、基因组图谱数据、来自分子模型数据库（MMDB）的蛋白质三维结构数据、种群序列

数据集以及由 PubMed 获得 Medline 的文献数据。Entrez 提供了方便实用的检索服务，所有操作都可以在网络浏览器上完成。用户可以利用 Entrez 界面上提供的限制条件（Limits）、索引（Index）、检索历史（History）和剪贴板（Clipboard）等功能来实现复杂的检索查询工作。对于检索获得的记录，用户可以选择需要显示的数据，保存查询结果，甚至以图形方式观看检索获得的序列。更详细的 Entrez 使用说明可以在该主页上获得。

2. 向 Genbank 提交序列数据　测序工作者可以把自己工作中获得的新序列提交给 NCBI，添加到 Genbank 数据库。这个任务可以由基于 Web 界面的 BankIt 或独立程序 Sequin 来完成。BankIt 是一系列表单，包括联络信息、发布要求、引用参考信息、序列来源信息以及序列本身的信息等。用户提交序列后，会从电子邮件收到自动生成的数据条目，Genbank 的新序列编号，以及完成注释后的完整的数据记录。用户还可以在 BankIt 页面下修改已经发布序列的信息。BankIt 适合于独立测序工作者提交少量序列，而不适合大量序列的提交，也不适合提交很长的序列，EST 序列和 GSS 序列也不应用 BankIt 提交。BankIt 使用说明和对序列的要求可详见其主页面。大量的序列提交可以由 Sequin 程序完成。Sequin 程序能方便的编辑和处理复杂注释，并包含一系列内建的检查函数来提高序列的质量保证。它还被设计用于提交来自系统进化、种群和突变研究的序列，可以加入比对的数据。Sequin 除了用于编辑和修改序列数据记录，还可以用于序列的分析，任何以 FASTA 或 ASN.1 格式序列为输入数据的序列分析程序都可以整合到 Sequin 程序下。在不同操作系统下运行的 Sequin 程序都可以在 ftp://ncbi.nlm.nih.gov/sequin/下找到，Sequin 的使用说明可详见其网页。

NCBI 的网址是：http://www.ncbi.nlm.nih.gov/。

Entrez 的网址是：http://www.ncbi.nlm.nih.gov/entrea/。

BankIt 的网址是：http://www.ncbi.nlm.nih.gov/BankIt。

Sequin 的相关网址是：http://www.ncbi.nlm.nih.gov/Sequin/。

（二）EMBL 核酸序列数据库

EMBL 核酸序列数据库由欧洲生物信息学研究所（EBI）维护的核酸序列数据构成，由于与 Genbank 和 DDBJ 的数据合作交换，它也是一个全面的核酸序列数据库。该数据库由 Oracal 数据库系统管理维护，查询检索可以通过通过因特网上的序列提取系统（SRS）服务完成。向 EMBL 核酸序列数据库提交序列可以通过基于 Web 的 WEBIN 工具，也可以用 Sequin 软件来完成。

数据库网址是：http://www.ebi.ac.uk/embl/。

SRS 的网址是：http://srs.ebi.ac.uk/。

WEBIN 的网址是：http://www.ebi.ac.uk/embl/Submission/webin.html。

（三）DDBJ 数据库

日本 DNA 数据仓库（DDBJ）也是一个全面的核酸序列数据库，与 Genbank 和 EMBL 核酸库合作交换数据。可以使用其主页上提供的 SRS 工具进行数据检索和序列

分析。可以用 Sequin 软件向该数据库提交序列。

DDBJ 的网址是：http：//www. ddbj. nig. ac. jp/。

二、蛋白质数据库

(一) PIR 和 PSD

PIR 国际蛋白质序列数据库（PSD）是由蛋白质信息资源（PIR）、慕尼黑蛋白质序列信息中心（MIPS）和日本国际蛋白质序列数据库（JIPID）共同维护的国际上最大的公共蛋白质序列数据库。这是一个全面的、经过注释的、非冗余的蛋白质序列数据库，包含超过 142000 条蛋白质序列（至 1999 年 9 月），其中包括来自几十个完整基因组的蛋白质序列。所有序列数据都经过整理，超过 99％的序列已按蛋白质家族分类，一半以上还按蛋白质超家族进行了分类。PSD 的注释中还包括对许多序列、结构、基因组和文献数据库的交叉索引，以及数据库内部条目之间的索引，这些内部索引帮助用户在包括复合物、酶－底物相互作用、活化和调控级联和具有共同特征的条目之间方便的检索。每季度都发行一次完整的数据库，每周可以得到更新部分。PSD 数据库有几个辅助数据库，如基于超家族的非冗余库等。PIR 提供三类序列搜索服务：基于文本的交互式检索；标准的序列相似性搜索，包括 BLAST，FASTA 等；结合序列相似性、注释信息和蛋白质家族信息的高级搜索，包括按注释分类的相似性搜索、结构域搜索 GeneFIND 等。

PIR 和 PSD 的网址是：http：//pir. georgetown. edu。

数据库下载地址是：ftp：//nbrfa. georgetown. edu/pir/。

(二) SWISS－PROT

SWISS－PROT 是经过注释的蛋白质序列数据库，由欧洲生物信息学研究所（EBI）维护。数据库由蛋白质序列条目构成，每个条目包含蛋白质序列、引用文献信息、分类学信息、注释等，注释中包括蛋白质的功能、转录后修饰、特殊位点和区域、二级结构、四级结构、与其他序列的相似性、序列残缺与疾病的关系、序列变异体和冲突等信息。SWISS －PROT 中尽可能减少了冗余序列，并与其他 30 多个数据建立了交叉引用，其中包括核酸序列库、蛋白质序列库和蛋白质结构库等。利用序列提取系统（SRS）可以方便地检索 SWISS －PROT 和其他 EBI 的数据库。SWISS－PROT 只接受直接测序获得的蛋白质序列，序列提交可以在其 Web 页面上完成。

SWISS－PROT 的网址是：http：//www. ebi. ac. uk/swissprot/。

(三) PROSITE

PROSITE 数据库收集了生物学有显著意义的蛋白质位点和序列模式，并能根据这些位点和模式快速和可靠地鉴别一个未知功能的蛋白质序列应该属于哪一个蛋白质家族。有的情况下，某个蛋白质与已知功能蛋白质的整体序列相似性很低，但由于功能的需要保留了与功能密切相关的序列模式，这样就可能通过 PROSITE 的搜索找到隐含的

功能模序，因此是序列分析的有效工具。PROSITE 中涉及的序列模式包括酶的催化位点、配体结合位点、与金属离子结合的残基、二硫键的半胱氨酸、与小分子或其他蛋白质结合的区域等；除了序列模式之外，PROSITE 还包括由多序列比对构建的 profile，能更敏感地发现序列与 profile 的相似性。PROSITE 的主页上提供各种相关检索服务。

PROSITE 的网址：http://www.expasy.ch/prosite/。

（四）PDB

蛋白质数据仓库（PDB）是国际上唯一的生物大分子结构数据档案库，由美国 Brookhaven 国家实验室建立。PDB 收集的数据来源于 X 光晶体衍射和核磁共振（NMR）的数据，经过整理和确认后存档而成。目前 PDB 数据库的维护由结构生物信息学研究合作组织（RCSB）负责。RCSB 的主服务器和世界各地的镜像服务器提供数据库的检索和下载服务，以及关于 PDB 数据文件格式和其他文档的说明，PDB 数据还可以从发行的光盘获得。使用 Rasmol 等软件可以在计算机上按 PDB 文件显示生物大分子的三维结构。

RCSB 的 PDB 数据库网址是：http://www.rcsb.org/pdb/。

（五）SCOP

蛋白质结构分类（SCOP）数据库详细描述了已知的蛋白质结构之间的关系。分类基于若干层次：家族，描述相近的进化关系；超家族，描述远源的进化关系；折叠子（fold），描述空间几何结构的关系；折叠类，所有折叠子被归于全 α，全 β，α/β，α+β 和多结构域等几个大类。SCOP 还提供一个非冗余的 ASTRAIL 序列库，这个库通常被用来评估各种序列比对算法。此外，SCOP 还提供一个 PDB－ISL 中介序列库，通过与这个库中序列的两两比对，可以找到与未知结构序列远缘的已知结构序列。

SCOP 的网址是：http://scop.mrc-lmb.cam.ac.uk/scop/。

（六）COG

蛋白质直系同源簇（COGs）数据库是对细菌、藻类和真核生物的 21 个完整基因组的编码蛋白，根据系统进化关系分类构建而成。COG 库对于预测单个蛋白质的功能和整个新基因组中蛋白质的功能都很有用。利用 COGNITOR 程序，可以把某个蛋白质与所有 COGs 中的蛋白质进行比对，并把它归入适当的 COG 簇。COG 库提供了对 COG 分类数据的检索和查询，基于 Web 的 COGNITOR 服务，系统进化模式的查询服务等。

COG 库的网址是：http://www.ncbi.nlm.nih.gov/COG。

下载 COG 库和 COGNITOR 程序在：ftp://ncbi.nlm.nih.gov/pub/COG。

三、功能数据库

（一）KEGG

京都基因和基因组百科全书（KEGG）是系统分析基因功能，联系基因组信息和功

能信息的知识库。基因组信息存储在 GENES 数据库里，包括完整和部分测序的基因组序列；更高级的功能信息存储在 PATHWAY 数据库里，包括图解的细胞生化过程如代谢、膜转运、信号传递、细胞周期，还包括同系保守的子通路等信息；KEGG 的另一个数据库是 LIGAND，包含关于化学物质、酶分子、酶反应等信息。KEGG 提供了 Java 的图形工具来访问基因组图谱，比较基因组图谱和操作表达图谱，以及其他序列比较、图形比较和通路计算的工具，可以免费获取。

KEGG 的网址是：http：//www. genome. ad. jp/kegg/。

（二）DIP

相互作用的蛋白质数据库（DIP）收集了由实验验证的蛋白质－蛋白质相互作用。数据库包括蛋白质的信息、相互作用的信息和检测相互作用的实验技术三个部分。用户可以根据蛋白质、生物物种、蛋白质超家族、关键词、实验技术或引用文献来查询 DIP 数据库。

DIP 的网址是：http：//dip. doe－mbi. ucla. edu/。

（三）ASDB

可变剪接数据库（ASDB）包括蛋白质库和核酸库两部分。ASDB（蛋白质）部分来源于 SWISS－PROT 蛋白质序列库，通过选取有可变剪接注释的序列，搜索相关可变剪接的序列，经过序列比对、筛选和分类构建而成。ASDB（核酸）部分来自 Genbank 中提及和注释的可变剪接的完整基因构成。数据库提供了方便的搜索服务。

ASDB 的网址是：http：//cbcg. nersc. gov/asdb。

（四）TRRD

转录调控区数据库（TRRD）是在不断积累的真核生物基因调控区结构－功能特性信息基础上构建的。每一个 TRRD 的条目里包含特定基因各种结构－功能特性：转录因子结合位点、启动子、增强子、静默子以及基因表达调控模式等。TRRD 包括五个相关的数据表：TRRDGENES（包含所有 TRRD 库基因的基本信息和调控单元信息）；TRRDSITES（包括调控因子结合位点的具体信息）；TRRDFACTORS（包括 TRRD 中与各个位点结合的调控因子的具体信息）；TRRDEXP（包括对基因表达模式的具体描述）；TRRDBIB（包括所有注释涉及的参考文献）。TRRD 主页提供了对这几个数据表的检索服务。

TRRD 的网址是：http：//wwwmgs. bionet. nsc. ru/mgs/dbases/trrd4/。

（五）TRANSFAC

TRANSFAC 数据库是关于转录因子、它们在基因组上的结合位点和与 DNA 结合的 profiles 的数据库。由 SITE，GENE，FACTOR，CLASS，MATRIX，CELLS，METHOD 和 REFERENCE 等数据表构成。此外，还有几个与 TRANSFAC 密切相关的扩展库：PATHODB 库收集了可能导致病态的突变的转录因子和结合位点；S/

MART DB 收集了与染色体结构变化相关的蛋白因子和位点的信息；TRANSPATH 库用于描述与转录因子调控相关的信号传递的网络；CYTOMER 库表现了人类转录因子在各个器官、细胞类型、生理系统和发育时期的表达状况。TRANSFAC 及其相关数据库可以免费下载，也可以通过 Web 进行检索和查询。

TRANSFAC 的网址是：http：//transfac. gbf. de/TRANSFAC/。

除了以上提及的数据库之外，还有许许多多的专门生物信息数据库，涉及了目前生物学研究的各个层面和领域，由于篇幅所限无法一一详述。国内也有一些大数据库的镜像站点和自己开发的有特色的数据库，如欧洲分子生物学网络组织 EMBNet 中国节点北京大学分子生物信息镜像系统，上海博容基因公司与上海嘉瑞软件公司合作开发的国产汉化基因数据库及分析管理系统，同时国家级的生物信息学中心也在筹建之中。

清华大学生物信息学研究所网址：http：//bioinfo. tsinghua. edu. cn/。

北京大学生物信息镜像系统网址：http：//cbi. pku. edu. cn/。

四、序列比对和预测分析

比较是科学研究中最常见的方法，通过将研究对象相互比较来寻找对象可能具备的特性。在生物信息学研究中，比对是最常用和最经典的研究手段。最常见的比对是蛋白质序列之间或核酸序列之间的两两比对，通过比较两个序列之间的相似区域和保守性位点，寻找二者可能的分子进化关系。进一步的比对是将多个蛋白质或核酸同时进行比较，寻找这些有进化关系的序列之间共同的保守区域、位点和 profile，从而探索导致它们产生共同功能的序列模式。此外，还可以把蛋白质序列与核酸序列相比来探索核酸序列可能的表达框架；把蛋白质序列与具有三维结构信息的蛋白质相比较，从而获得蛋白质折叠类型的信息。比对还是数据库搜索算法的基础，将查询序列与整个数据库的所有序列进行比对，从数据库中获得与其最相似序列的已有的数据，能最快速地获得有关查询序列的大量有价值的参考信息，对于进一步分析其结构和功能都会有很大的帮助。近年来随着生物信息学数据大量积累和生物学知识的整理，通过比对方法可以有效地分析和预测一些新发现基因的功能。

（一）序列两两比对

序列比对的理论基础是进化学说，如果两个序列之间具有足够的相似性，就推测二者可能有共同的进化祖先，经过序列内残基的替换、残基或序列片段的缺失、序列重组等遗传变异过程分别演化而来。序列相似和序列同源是不同的概念，序列之间的相似程度是可以量化的参数，而序列是否同源需要有进化事实的验证。在残基－残基比对中，可以明显看到序列中某些氨基酸残基比其他位置上的残基更保守，这些信息揭示了这些保守位点上的残基对蛋白质的结构和功能是至关重要的，例如它们可能是酶的活性位点残基，形成二硫键的半胱氨酸残基，与配体结合部位的残基，与金属离子结合的残基，形成特定结构 motif 的残基等等。但并不是所有保守的残基都一定是结构功能重要的，可能它们只是由于历史的原因被保留下来，而不是由于进化压力而保留下来。因此，如果两个序列有显著的保守性，要确定二者具有共同的进化历史，进面认为二者有近似的

结构和功能还需要更多实验和信息的支持。通过大量实验和序列比对的分析，一般认为蛋白质的结构和功能比序列具有更大的保守性，因此粗略的说，如果序列之间的相似性超过30%，它们就很可能是同源的。

（二）核酸与蛋白质结构和功能的预测分析

人们获得各种核酸和蛋白质序列的目的是了解这个序列在生物体中充当了怎样的角色。例如，DNA序列中重复片段、编码区、启动子、内含子/外显子、转录调控因子结合位点等信息；蛋白质的分子量、等电点、二级结构、三级结构、四级结构、膜蛋白的跨膜区段、酶的活性位点、蛋白质之间相互作用等结构和功能信息。虽然用实验的方法是多年以来解决这类问题的主要途径，但新的思路是利用已有的对生物大分子结构和功能特性的认识，用生物信息学的方法通过计算机模拟和计算来“预测”出这些信息或提供与之相关的辅助信息。由于生物信息学的特点，可以用较低的成本和较快的时间就能获得可靠的结果。近10年来生物学序列信息的爆炸性增长大大促进了各种序列分析和预测技术的发展，目前已经可以用理论预测的方法获得大量的结构和功能信息。要注意的是，尽管各种预测方法都基于现有的生物学数据和已有的生物学知识，但在不同模型或算法基础上建立的不同分析程序有其一定的适用范围和相应的限制条件，因此最好对同一个生物学问题尽量多用几种分析程序，综合分析各种方法得到的结果和结果的可靠性。此外，生物信息学的分析只是为生物学研究提供参考，这些信息能提高研究的效率或提供研究的思路，但很多问题还需要通过实验的方法得到验证。

1. 针对核酸序列的预测方法　针对核酸序列的预测就是在核酸序列中寻找基因，找出基因的位置和功能位点的位置，以及标记已知的序列模式等过程。在此过程中，确认一段DNA序列是一个基因需要有多个证据的支持。一般而言，在重复片段频繁出现的区域里，基因编码区和调控区不太可能出现；如果某段DNA片段的假想产物与某个已知的蛋白质或其他基因的产物具有较高序列相似性的话，那么这个DNA片段就非常可能属于外显子片段；在一段DNA序列上出现统计上的规律性，即所谓的“密码子偏好性”，也是说明这段DNA是蛋白质编码区的有力证据；其他的证据包括与“模板”序列的模式相匹配、简单序列模式如TATA Box等相匹配等。一般而言，确定基因的位置和结构需要多个方法综合运用，而且需要遵循一定的规则：对于真核生物序列，在进行预测之前先要进行重复序列分析，把重复序列标记出来并除去；选用预测程序时要注意程序的物种特异性；要弄清程序适用的是基因组序列还是cDNA序列；很多程序对序列长度也有要求，有的程序只适用于长序列，而对EST这类残缺的序列则不适用。

2. 针对蛋白质的预测方法　传统的生物学认为，蛋白质的序列决定了它的三维结构，也就决定了它的功能。由于用X光晶体衍射和NMR核磁共振技术测定蛋白质的三维结构，以及用生化方法研究蛋白质的功能效率不高，无法适应蛋白质序列数量飞速增长的需要，因此近几十年来许多科学家致力于研究用理论计算的方法预测蛋白质的三维结构和功能，经过多年努力取得了一定的成果。

3. 从氨基酸组成辨识蛋白质　根据组成蛋白质的20种氨基酸的物理和化学性质可以分析电泳等实验中的未知蛋白质，也可以分析已知蛋白质的物化性质。ExPASy工具

包中提供了一系列相应程序：

AACompIdent：根据氨基酸组成辨识蛋白质。这个程序需要的信息包括：氨基酸组成、蛋白质的名称（在结果中有用）、pI 和 MW（如果已知）以及它们的估算误差、所属物种或物种种类或“全部（ALL）”、标准蛋白的氨基酸组成、标准蛋白的 SWISS－PROT 编号、用户的 Email 地址等，其中一些信息可以没有。这个程序在 SWISS－PROT 和 TrEMBL 数据库中搜索组成相似蛋白。

AACompSim：与前者类似，但比较在 SWISS－PROT 条目之间进行。这个程序可以用于发现蛋白质之间较弱的相似关系。

除了 ExPASy 中的工具外，PROPSEARCH 也提供基于氨基酸组成的蛋白质辨识功能。程序作者用 144 种不同的物化性质来分析蛋白质，包括分子量、巨大残基的含量、平均疏水性、平均电荷等，把查询序列的这些属性构成的“查询向量”与 SWISS－PROT 和 PIR 中预先计算好的各个已知蛋白质的属性向量进行比较。这个工具能有效地发现同一蛋白质家族的成员。可以通过 Web 使用这个工具，用户只需输入查询序列本身。

ExPASy 的网址是：http：//www. expasy. ch/tools/。

PROSEARCH 的网址是：http：//www. embl－heidelberg. de/prs. html。

4. 预测蛋白质的物理性质　从蛋白质序列出发，可以预测出蛋白质的许多物理性质，包括等电点、分子量、酶切特性、疏水性、电荷分布等。相关工具有：

Compute pI/MW：是 ExPASy 工具包中的程序，计算蛋白质的等电点和分子量。对于碱性蛋白质，计算出的等电点可能不准确。

PeptideMass：是 ExPASy 工具包中的程序，分析蛋白质在各种蛋白酶和化学试剂处理后的内切产物。蛋白酶和化学试剂包括胰蛋白酶、糜蛋白酶、LysC、溴化氰、ArgC，AspN 和 GluC 等。

TGREASE：是 FASTA 工具包中的程序，分析蛋白质序列的疏水性。这个程序按序列计算每个残基位点的移动平均疏水性，并给出疏水性－序列曲线，用这个程序可以发现膜蛋白的跨膜区和高疏水性区的明显相关性。

SAPS：蛋白质序列统计分析，对提交的序列给出大量全面的分析数据，包括氨基酸组成统计、电荷分布分析、电荷聚集区域、高度疏水区域、跨膜区段等等。

下载 FASTA 的网址是：ftp：//ftp. virginia. edu/pub/fasta/。

SAPS 的网址是：http：//www. isrec. isb－sib. ch/software/SAPS form. html。

5. 蛋白质二级结构预测　二级结构是指 α 螺旋和 β 折叠等规则的蛋白质局部结构元件。不同的氨基酸残基对子形成不同的二级结构元件具有不同的倾向性。按蛋白质中二级结构的成分可以把球形蛋白分为全 α 蛋白、全 β 蛋白、α＋β 蛋白和 α/β 蛋白等四个折叠类型。预测蛋白质二级结构的算法大多以已知三维结构和二级结构的蛋白质为依据，用人工神经网络、遗传算法等技术构建预测方法。还有将多种预测方法结合起来，获得“一致序列”。总的来说，二级结构预测仍是未能完全解决的问题，一般对于 α 螺旋预测精度较好，对 β 折叠差些，面对除 α 螺旋和 β 折叠等之外的无规则二级结构则效果很差。

nnPredict：用神经网络方法预测二级结构，蛋白质结构类型分为全α蛋白、全β蛋白和α/β蛋白，输出结果包括“H”（螺旋）、“E”（折叠）和“—”（转角）。这个方法对全α蛋白能达到79%的准确率。

PredictProtein：提供了序列搜索和结构预测服务。它先在SWISS－PROT中搜索相似序列，用MaxHom算法构建多序列比对的profile，再在数据库中搜索相似的profile，然后用一套PHD程序来预测相应的结构特征，包括二级结构。返回的结果包含大量预测过程中产生的信息，还包含每个残基位点的预测可信度。这个方法的平均预测准确率达到72%。

SOPMA：带比对的自优化预测方法，将几种独立二级结构预测方法汇集成“一致预测结果”，采用的二级结构预测方法包括GOR方法、Levin同源预测方法、双重预测方法、PHD方法和SOPMA方法。多种方法的综合应用平均效果比单个方法更好。

nnPredict的网址是：http：//www.cmpharm.ucsf.edu/～nomi/nnpredict.html。

PredictProtein的网址是：http：//cubic.bioc.columbia.edu/predictprotein。

PredictProtein的国内镜像在：http：//www.cbi.pku.edu.cn/predictprotein/。

SOPMA的网址是：http：//pbil.ibcp.fr/。

6. 其他特殊局部结构　其他特殊局部结构包括膜蛋白的跨膜螺旋、信号肽、卷曲螺旋（Coiled Coils）等，具有明显的序列特征和结构特征，也可以用计算方法加以预测。

COILS：卷曲螺旋预测方法，将序列与已知的平行双链卷曲螺旋数据库进行比较，得到相似性得分，并据此算出序列形成卷曲螺旋的概率。

TMpred：预测蛋白质的跨膜区段和在膜上的取向，它根据来自SWISS－PROT的跨膜蛋白数据库Tmbase，利用跨膜结构区段的数量、位置以及侧翼信息，通过加权打分进行预测。

SignalP：预测蛋白质序列中信号肽的剪切位点。

COILS的网址是：http：//www.ch.embnet.org/software/COILS form.html。

TMpred的网址是：http：//www.ch.embnet.org/software/TMPRED form.html。

SignalP的网址是：http：//www.cbs.dtu.dk/services/SignalP/。

（7）蛋白质的三维结构　蛋白质三维结构预测是最复杂和最困难的预测技术。研究发现，序列差异较大的蛋白质也可能折叠成类似的三维构象，自然界里的蛋白质结构骨架的多样性远少于蛋白质序列的多样性。由于蛋白质的折叠过程仍然不十分明了，从理论上解决蛋白质折叠的问题还有待进一步的科学发展，但也有了一些有一定作用的三维结构预测方法。最常见的是“同源模建”和“Threading”方法。前者先在蛋白质结构数据库中寻找未知结构蛋白的同源伙伴，再利用一定计算方法把同源蛋白的结构优化构建出预测的结果。后者将序列“穿”入已知的各种蛋白质的折叠子骨架内，计算出未知结构序列折叠成各种已知折叠子的可能性，由此可预测序列分配最合适的折叠子结构。除了“Threading”方法外，用PSI－BLAST方法也可以把查询序列分配到合适的蛋白质折叠家族，实际应用中发现该方法的效果也不错。

SWISS－MODEL：自动蛋白质同源模建服务器，有两个工作模式：第一步模式（First Approach mode）和优化模式（Optimise mode）。程序先把提交的序列在 ExPdb 晶体图像数据库中搜索相似性足够高的同源序列，建立最初的原子模型，再对这个模型进行优化产生预测的结构模型。

CPHmodels：也是利用神经网络进行同源模建预测蛋白质结构的方法。

SWISS－MODEL 的网址是：http：//www. expasy. ch/swissmod/SWISS－MODEL. html。

CPHmodels 的网址是：http：//www. cbs. dtu. dk/services/CPHmodels/。

第三节 生物信息学在蛋白质研究中的应用

一、蛋白质组

基因组对生命体的整体控制必须通过它所表达的全部蛋白质来执行，由于基因芯片技术只能反映从基因组到 RNA 的转录水平上的表达情况，而从 RNA 到蛋白质还有许多中间环节的影响，因此仅凭基因芯片技术我们还不能最终掌握生物功能具体执行者——蛋白质的整体表达状况，因此，近几年在发展基因芯片的同时，人们也发展了一套研究基因组所有蛋白质产物表达情况——蛋白质组研究技术，从技术上来讲包括二维凝胶电泳技术和质谱测序技术。通过二维凝胶电泳技术可以获得某一时间截面上蛋白质组的表达情况，通过质谱测序技术就可以得到所有这些蛋白质的序列组成。这些都是技术实现问题，最重要的就是如何运用生物信息学理论方法去分析所得到的巨量数据，从中还原出生命运转和调控的整体系统的分子机制。

二、蛋白质结构

基因组和蛋白质组研究的迅猛发展，使许多新蛋白序列涌现出来，然而要想了解它们的功能，只有氨基酸序列是远远不够的，因为蛋白质的功能是通过其三维高级结构来执行的，而且蛋白质三维结构也不一定是静态的，在行使功能的过程中其结构也会相应的有所改变。因此，得到这些新蛋白的完整、精确和动态的三维结构就成为摆在我们面前的紧迫任务。目前除了通过诸如 X 射线晶体结构分析、多维核磁共振（NMR）波谱分析和电子显微镜二维晶体三维重构（电子晶体学，EC）等物理方法得到蛋白质三维结构之外，另外一种广泛使用的方法就是通过计算机辅助预测的方法，目前，一般认为蛋白质的折叠类型只有数百到数千种，远远小于蛋白质所具有的自由度数目，而且蛋白质的折叠类型与其氨基酸序列具有相关性，这样就有可能直接从蛋白质的氨基酸序列通过计算机辅助方法预测出蛋白质的三维结构。

三、新药设计

近年来随着结构生物学的发展，相当数量的蛋白质以及一些核酸、多糖的三维结构获得精确测定，基于生物大分子结构知识的药物设计成为当前的热点。生物信息学的研

究不仅可提供生物大分子空间结构的信息，还能提供电子结构的信息，如能级、表面电荷分布、分子轨道相互作用等以及动力学行为的信息，如生物化学反应中的能量变化、电荷转移、构象变化等。理论模拟还可研究包括生物分子及其周围环境的复杂体系和生物分子的量子效应。通过以上工作可为新药设计提供重要信息。

（王晓明　刘师蓬）

主要参考文献

1. 吴乃虎主编．基因工程学原理（上，下册）．北京：科学出版社，2000
2. 陈惠黎主编．生物大分子的结构和功能．上海：上海医科大学出版社，1999
3. 周爱儒主编．生物化学（第五版）．北京：人民卫生出版社，2000
4. 冯作化主编．医学分子生物学．北京：人民卫生出版社，2001
5. 孙树汉主编．基因工程原理与方法．北京：人民军医出版社，2001
6. 郭葆玉主编．基因工程药学．上海：第二军医大学出版社，2000
7. 徐松德、张建业主编．医用分子生物学．北京：人民卫生出版社，1999
8. 张龙翔、张庭芳、李令媛主编．生化实验方法和技术（第二版），北京：高等教育出版社，1997
9. 于世林编著．高效液相色谱方法及应用．化学工业出版社，2000
10. 张龙翔．丝氨酸蛋白酶的蛋白质工程．生命的化学，1992；12（2）：6～11
11. 鲁子贤．蛋白质工程研究的推敲．生命的化学；1990；10（6）：12～14
12. 夏其昌主编．蛋白质化学研究技术与进展．北京：科技出版社，1997
13. 徐秀璋．蛋白质顺序分析技术．北京：科学出版社，1988
14. 谢忠信．X 射线光谱分析．北京：科学出版社，1982
15. 王金山．核磁共振波谱仪与实验技术．北京：机械工业出版社，1982
16. 张红缨等．蛋白质工程的新策略—酶的体外定向进化．科学通报，1999；29：
17. 朱国萍等．葡萄糖异构酶的生物工程研究进展．生物化学与生物物理进展，2000；27（2）：127～131
18. 王丸强等．枯草杆菌蛋白酶基因工程的研究进展．生物工程进展，2000；20（2）：41～44
19. 王琼庆、冯佶民．胰岛素蛋白质工程研究进展．生物化学与生物物理进展，1996；23：402～408
20. 朱国萍等．G138P 定点突变对葡萄糖异构酶稳定性的改善．生物化学与生物物理学报，1998；607～610
21. 黄志伟、王琰．抗体工程．北京．北京医科大学和中国协和医科大学联合出版社．1997
22. Robert F. Weaver. 分子生物学（Molecular Biology）．北京：科学出版

社，2000

23. Trudy Mokee, James R. Mckee. 生物化学导论（Biochemistry An Introduction 2nd Ed）. 北京：科学出版社。
24. Bernard R. Glick, Jack J. Pasternak. Molecular Biotechnology: Principle and Application of Recombinant DNA. Washington, D. C: ASM Press 1994
25. Ulmer K. M. Protein Engineering. Science, 1983; 219: 666～671
26. Robson B, Garnier J. Introduction to Protein Engineering. Elsevier, 1986
27. Shevchenko A, Jensen DN, Podtelejnikov AV, et al. Linking genome and proteome by mass spectrometry: large－scale identification of yeast proteins from two dimensional gels [J]. PNAS, 1996, 93 (25): 14440
28. Welnam KJ, Domin MA, Scannell DE, et al. The characterization of microorganisms by matrix－assisted laser deorption/ionization time－of－flight mass spectrometry [J]. Rapid Commun Mass Spectrom, 1998, 12 (4): 176
29. Clore, G. M. and Gronenborn, A. M, In NMR of Proteins, Macmillan Press Ltd, 1993

社，2000

23. Trudy Mokee, James R. Mckee. 生物化学导论（Biochemistry An Introduction 2nd Ed）. 北京：科学出版社。
24. Bernard R. Glick, Jack J. Pasternak. Molecular Biotechnology: Principle and Application of Recombinant DNA. Washington, D. C: ASM Press 1994
25. Ulmer K. M. Protein Engineering. Science, 1983; 219: 666～671
26. Robson B, Garnier J. Introduction to Protein Engineering. Elsevier, 1986
27. Shevchenko A, Jensen DN, Podtelejnikov AV, et al. Linking genome and proteome by mass spectrometry: large－scale identification of yeast proteins from two dimensional gels [J]. PNAS, 1996, 93 (25): 14440
28. Welnam KJ, Domin MA, Scannell DE, et al. The characterization of microorganisms by matrix－assisted laser deorption/ionization time－of－flight mass spectrometry [J]. Rapid Commun Mass Spectrom, 1998, 12 (4): 176
29. Clore, G. M. and Gronenborn, A. M, In NMR of Proteins, Macmillan Press Ltd, 1993

社，2000

23. Trudy Mokee, James R. Mckee. 生物化学导论（Biochemistry An Introduction 2nd Ed）. 北京：科学出版社。
24. Bernard R. Glick, Jack J. Pasternak. Molecular Biotechnology: Principle and Application of Recombinant DNA. Washington, D. C: ASM Press 1994
25. Ulmer K. M. Protein Engineering. Science, 1983; 219: 666～671
26. Robson B, Garnier J. Introduction to Protein Engineering. Elsevier, 1986
27. Shevchenko A, Jensen DN, Podtelejnikov AV, et al. Linking genome and proteome by mass spectrometry: large－scale identification of yeast proteins from two dimensional gels [J]. PNAS, 1996, 93 (25): 14440
28. Welnam KJ, Domin MA, Scannell DE, et al. The characterization of microorganisms by matrix－assisted laser deorption/ionization time－of－flight mass spectrometry [J]. Rapid Commun Mass Spectrom, 1998, 12 (4): 176
29. Clore, G. M. and Gronenborn, A. M, In NMR of Proteins, Macmillan Press Ltd, 1993

社，2000

23. Trudy Mokee, James R. Mckee. 生物化学导论（Biochemistry An Introduction 2nd Ed）. 北京：科学出版社。
24. Bernard R. Glick, Jack J. Pasternak. Molecular Biotechnology: Principle and Application of Recombinant DNA. Washington, D. C: ASM Press 1994
25. Ulmer K. M. Protein Engineering. Science, 1983; 219: 666～671
26. Robson B, Garnier J. Introduction to Protein Engineering. Elsevier, 1986
27. Shevchenko A, Jensen DN, Podtelejnikov AV, et al. Linking genome and proteome by mass spectrometry: large－scale identification of yeast proteins from two dimensional gels [J]. PNAS, 1996, 93 (25): 14440
28. Welnam KJ, Domin MA, Scannell DE, et al. The characterization of microorganisms by matrix－assisted laser deorption/ionization time－of－flight mass spectrometry [J]. Rapid Commun Mass Spectrom, 1998, 12 (4): 176
29. Clore, G. M. and Gronenborn, A. M, In NMR of Proteins, Macmillan Press Ltd, 1993